Pattern Recognition Applied to Cough Categorization

By

JOSEPH L. DEVINE, JR.

A. J. WELCH

September 15, 1967

Technical Report No. 40

LABORATORIES FOR ELECTRONICS AND
RELATED SCIENCE RESEARCH

College of Engineering

THE UNIVERSITY OF TEXAS

AUSTIN, TEXAS 78712

12

The University of Texas' Laboratories for Electronics and Related Science Research are interdisciplinary laboratories in which graduate faculty members and graduate candidates from numerous academic disciplines conduct research.

Research conducted for this technical report was supported in part by the Department of Defense's JOINT SERVICES ELECTRONICS PROGRAM (U. S. Army, U. S. Navy, and the U. S. Air Force) through the Research Grant AF-AFOSR-776-67. This program is monitored by the Department of Defense's JSEP Technical Advisory Committee consisting of representatives from the U. S. Army Electronics Command, U. S. Army Research Office, Office of Naval Research, and the U. S. Air Force Office of Scientific Research.

Additional support of specific projects by other Federal Agencies, Foundations, and The University of Texas is acknowledged in footnotes to the appropriate sections.

Requests for additional copies by agencies of the Department of Defense, their contractors and other government agencies should be directed to:

> Defense Documentation Center (DDC)
> Cameron Station
> Alexandria, Virginia 22314

Department of Defense contractors must be established for DDC services to have their "need to know" certified by the cognizant military agency of their project or contract.

THE UNIVERSITY OF TEXAS

LABORATORIES FOR ELECTRONICS AND RELATED SCIENCE RESEARCH

PATTERN RECOGNITION APPLIED TO COUGH CATEGORIZATION*

Technical Report No. 40

By

Joseph L. Devine, Jr.**
Ashley J. Welch,***

September 15, 1967

* Research sponsored in part by the Joint Services Electronics Program under the Research Grant AF-AFOSR-766-67 and a grant from the Committee for Research on Tobacco and Health of the American Medical Association Education and Research Foundation.

** Graduate Student, Candidate for Ph.D. Degree in Electrical Engineering.

*** Assistant Professor of Electrical Engineering, The University of Texas.

ABSTRACT

The particular problem with which the research was concerned was the development of a technique to discriminate between coughs and other audible phenomena which originate in a hospital environment. Pattern recognition provided such a technique. Experimental data was available in the form of audio tape recordings.

Implementation of a Bayes categorization decision requires knowledge of the underlying conditional joint probability density functions of the measures which typify the patterns to be recognized. An adaptive pattern classifier model was presented which circumvented the difficulty of estimating these functions. The model is generally applicable to the two-class case in which the patterns to be classified consist of sequential segments of data known to have originated from the same class. The model took the form of a layered machine. The first stage was a minimum distance classifier with respect to point sets while the second stage utilized the first stage binary valued outputs to implement a Bayes decision.

The feature extraction and measure selection problems were examined experimentally. Feature calculation algorithms were developed which are generally applicable to time varying signals. The effectiveness of an algorithm for selection of a set of candidate measures was verified.

Experimental results indicated that, for the particular data with which the research was concerned, an assumption of Markoff-1 dependency between sequential first stage decisions of the pattern classifier was warranted. A pattern classifier which was based on this assumption classified 97.9% of the patterns presented to its input correctly.

TABLE OF CONTENTS

LIST OF TABLES

List of Tables cont'd.

FIGURE

CHAPTER I

INTRODUCTION

A group at The University of Texas Southwestern Medical School in Dallas, Texas is engaged in research in the general area of respiratory diseases.[*] One parameter of importance in their studies is the number of times that a patient coughs during a given time interval. The primary objective of the research outlined in this report is to develop a technique which may be mechanized, either as special purpose circuitry to operate in a real-time environment or as a general purpose computer program, to obtain this parameter.

The results may find additional application in pharmaceutical evaluation. Experience indicates that self-evaluation of cough activity by the patient is subjective to a pronounced degree. In the evaluation of a drug proposed as a cough alleviating agent, an objective method of evaluation of the results of administration of the drug is necessary.

For the purposes of this investigation audible phenomena which occur in the hospital environment, but which are not coughs, will be

---

[*] Research in progress under the supervision of Dr. R. G. Loudon, Department of Internal Medicine, The University of Texas Southwestern Medical School supported by a grant from the Committee for Research on Tobacco and Health of the American Medical Association Education and Research Foundation.

referred to as artifacts. A preliminary attempt by personnel at Southwestern Medical School to identify coughs by means of an empirically adjusted filter and voice controlled relay were unsuccessful. When adjusted for actuation by coughs, the system also responds to many artifacts. This indicated that more powerful classifying techniques would be required.

The purpose of the outlined research was to develop a means for discriminating between coughs and artifacts. Pattern recognition provides such a technique.

During recent years a good deal of effort has been directed toward the study of adpative pattern recognizers. These studies fall into two broad categories. The bulk of the literature in the field consists of mathematical studies indicating that a given procedure is optimal, or converges to an optimal procedure for a specified constraint under a given set of presumptions regarding underlying probability density functions. The second approach that is taken is to apply pattern recognition techniques to a set of data originating from a particular experiment for which the underlying probability functions are unknown or cannot be approximated by a computationally managable analytical expression and for which optimality cannot therefore be shown. The research outlined in this investigation falls into the latter category. The basics of pattern recognition are outlined in Chapter 2 of the report with emphasis placed on those portions of the theory that

bear on the particular pattern recognition machine selected for the described research. Data acquisition techniques and processing used to acquire the measures utilized in the recognition process and the basis on which the particular meaures chosen were selected are outlined in Chapter 3. The machine chosen to implement the decision process is described in Chapter 4. Experimental results are presented in Chapter 5.

CHAPTER II

GENERAL PATTERN RECOGNITION MODEL

2.1 Basic Model

A general review of pertinent portions of pattern recognition theory are presented in this chapter of the report. The notation employed is, for the most part, that used by Nillson.[10]

A pattern classifier has as an input a set of ordered (for our purposes, real) numbers gleaned from an experiment. These numbers will be referred to as measures and quantitatively describe an event. It will be convenient to visualize a set of d measures as a point in d--dimensional Euclidian space E^d. This space will be referred to as "pattern space". The rectangular coordinates in the point are the real numbers $x_1, x_2, \ldots x_d$. The vector $\underline{X}$ extending from the origin to the point $(x_1, x_2, \ldots x_d)$ will also be used to represent the pattern. $\underline{X}$ will be used to designate both the point and the vector. For computational purposes it will be considered a column matrix.

The output of the pattern classifier is a decision as to which of R classes the event described by the measures belongs, $R \geq 2$. The pattern recognizer, then, is a device which maps the points of E^d into the category numbers 1, 2, ..., R. Presuming points in E^d which map into different categories occupy disjoint regions in E^d, one may visualize

a partitioning of the space by surfaces arranged so that they separate the point sets belonging to the different categories. These are called decision surfaces. If the point sets are not disjoint, decision surfaces may still be placed in accordance with some predetermined criteria.

A pattern classifier is "adaptive" if it has the capability of modifying its performance (repositioning the decision surfaces) in accordance with measures presented to its input. The process of modification of the performance of the machine is termed "learning" while the act of aiding this learning is called "training". A set of data, chosen as typical, used to accomplish the training is called a "training set".

The manner in which training is conducted gives rise to two broad classifications of learning called "learning with a teacher" and "learning without a teacher". Learning without a teacher implies that the machine modifies its performance without explicit instruction as to the class membership of the training data. In learning with a teacher the true membership of the input measures is known a priori.

For the former type of machine, training may take place throughout the working cycle of the machine. In the latter case, training is done prior to the use of the machine as a classifier. The type of machine with which this report is concerned utilizes learning with a teacher.

2.2 Discriminant Functions

The decision surfaces previously described may be implicitly defined by a set of functions containing R members. Define $g_1(\underline{X})$, $g_2(\underline{X})$, ..., $g_R(\underline{X})$

as scalar, single-valued functions of the pattern $\underline{X}$ with the property that:

$$g_i(\underline{X}) > g_j(\underline{X}) \quad i, j = 1, 2, \ldots, R \quad (i \neq j) \quad \text{for all } \underline{X} \in \text{category } i. \tag{2-1}$$

These functions are called "discriminant functions." Presuming that the functions are continuous at the boundaries, it is seen that the surface separating contiguous regions i and j is defined by:

$$g_i(\underline{X}) - g_j(\underline{X}) = 0 \tag{2-2}$$

Discriminant functions provide a means for convenient implementation of pattern classification. For a pattern $\underline{X}$ which is to be classified as belonging to one of R classes, the R discriminant functions are calculated and the pattern is assigned to the category for which the associated discriminant function has the largest value.

A special case of particular interest is that where R = 2. In this case the partition of pattern space into the two regions is called a dichotomy. The associated discriminant functions may be combined to yield:

$$\begin{aligned} g(\underline{X}) &\stackrel{\Delta}{=} g_1(\underline{X}) - g_2(\underline{X}) \\ g(\underline{X}) &> 0 \Rightarrow \underline{X} \in \text{class } 1 \\ g(\underline{X}) &< 0 \Rightarrow \underline{X} \in \text{class } 2 \\ g(\underline{X}) &= 0 \Rightarrow \text{undefined or arbitrary classification} \end{aligned} \tag{2-3}$$

Discriminant functions may be selected in a variety of ways depending primarily upon the extent of a priori knowledge of the patterns to be classified. As previously indicated, the method employed in the described research utilizes training.

2.3 Training

Training methods have been broadly grouped into two categories--parametric and non-parametric.

The criteria for differentation between these two types of training are not consistent among all researchers in the field. Nilsson[10], for example, classifies as parametric any training which assumes that the patterns to be categorized are known a priori to be characterized by a set of parameters, some of which may be unknown. Sebestyen[12], on the other hand, reserves the parametric label for patterns originating from a random process with an underlying probability distribution function of known form which is described by a set of parameters, some of which may be unknown and which are themselves random variables. This latter definition is an extension of that encountered in estimation theory.[6]

In either case, training consists of computing estimates of the unknown parameters.

An important example of non-parametric pattern classification is that where the pattern space is deterministic in nature.

For the purpose of this report, training procedures will be labeled as probabilistic or non-probabilistic depending upon the assumptions made as to the nature of the pattern space. In some cases the line will not be clearly drawn.

2.4 Machine Structure

2.4.1 Statistical Decision Theory

If the pattern space is probabilistic and the underlying probability density functions and a priori probability of occurrences of the various classes are known, a decision criterion may be derived which is optimal under a given constraint.

Define:

$c(i/j)$ = cost of deciding $\underline{X} \epsilon i$ when in actuality $\underline{X} \epsilon j$.

$p(j/\underline{X})$ = conditional probability that $\underline{X} \epsilon j$ given that $\underline{X}$ has occurred.

$$L_{\underline{X}}(i) = \sum_{j=i}^{R} c(i/j)p(j/\underline{X}) = \text{conditional average loss for the decision } X \epsilon i; \text{ defined a priori.} \tag{2-4}$$

It is to be noted that $\underline{X}$ is a vector. $p(j/\underline{X})$ will therefore take the form of a conditional joint probability:

$$p(j/\underline{X}) = p(j/x_1, x_2, \ldots, x_d) \tag{2-5}$$

A decision which minimizes the above conditional average loss is called optimum and a machine which implements such a decision is called a Bayes machine.

Examining equation (2-4), it is seen that in order to derive an optimum decision a set of costs must be assigned which weight the relative importance of the different errors in classification. In addition, the a posteriori probabilities $p(j/\underline{X})$, $j = 1, 2, \ldots, R$ must be known. This

latter requirement is almost never accomplished in practice and would be difficult to estimate directly from experimental observations. Bayes' theorem, however, allows the calculation of the a posteriori probabilities in terms of estimable quantities:

$$p(j/\underline{X}) = \frac{p(\underline{X}/j)p(j)}{p(\underline{X})} \tag{2-6}$$

$p(\underline{X}/j)$ is called the "likelihood" of j with respect to $\underline{X}$. Substituting (2-6) into (2-4):

$$L_{\underline{X}}(j) = \frac{1}{p(\underline{X})} \sum_{j=1}^{R} c(i/j)p(\underline{X}/j)p(j) \tag{2-7}$$

It is seen that $1/p(\underline{X})$ is a common term for all j. Minimization of (2-7) is equivalent to minimization of (2-8):

$$\ell_{\underline{X}}(j) = \sum_{j=1}^{R} c(i/j)p(\underline{X}/j)p(j) \tag{2-8}$$

Equation (2-8) takes a form particularly amenable to calculation if the costs of incorrect decisions are made equal while the cost of a correct decision is set at zero:

$$c(i/j) = 1 - \delta_{ij} \tag{2-9}$$

where δ_{ij} is the Kronecker delta function.

Substitution of (2-9) into (2-8) yields:

$$\begin{aligned} \ell_{\underline{X}}(i) &= \sum_{j=1}^{R} p(\underline{X}/j)p(j) - p(\underline{X}/i)p(i) \\ &= p(\underline{X}) - p(\underline{X}/i)p(i) \end{aligned} \tag{2-10}$$

Minimization of (2-10) over all i, $i = 1, 2, \ldots, R$ implies maximization of $p(\underline{X}/i)p(i)$, $i = 1, 2, \ldots, R$. The discriminant functions for the R categories would therefore be the quantities $p(\underline{X}/i)p(i)$, $i = 1, 2, \ldots, R$, and the decision made would be in accordance with discriminant criteria outlined in section 2.2. For this special loss function, called a symmetrical loss function, the decision criterion has been given the name "Ideal Observer" criterion in decision theory. It can be shown that this decision rule minimizes the probability of erroneous classification.

In the event that the a priori probabilities $p(i)$, $i = 1, 2, \ldots, R$ are unknown and are therefore taken as being equally probable, the discriminant function is of the form $g_i(\underline{X}) = p(\underline{X}/i)$ and the criterion is known as the "Maximum Likelihood" criterion in decision theory. This criterion implies that $\underline{X}$ originates from the category for which its occurrence is most probable.

For the probabilistic decision criteria outlined, training consists of estimating the various probabilities. Since $\underline{X}$ is a vector, the probabilities $p(\underline{X}/i)$ are joint conditional probabilities and must be estimated for all points in pattern space. This is a formidable task if the dimensionality of $\underline{X}$ is large unless some simplifying assumptions are made. A common assumption that is made is that the underlying probability density functions are Gaussian, in which case training consists of estimating the unknown mean and covariance matrices of the density functions. This line of attack

will not be pursued since the measures employed do not exhibit Gaussian properties and any analytical calculations employing this assumption would be simply a mathematical exercise.

A brief summary of the probabilistic pattern recognition model for the symmetrical loss function is in order at this point. The input to the pattern classifier is a set of measures which is represented as a vector $\underline{X}$ in E^d. Training consists of estimating the associated likelihoods and a priori probabilities for all points in E^d for categories $i = 1, 2, \ldots,$ R. The discriminant function for category i is $g_i(\underline{X}) = p(\underline{X}/i)p(i)$. Classification of a vector $\underline{X}$ of unknown class membership is accomplished by evaluating $g_i(\underline{X})$ for $i = 1, 2, \ldots, R$ and assigning $\underline{X}$ to the category for which the associated $g(\underline{X})$ is largest. Decision surfaces are defined by the relationshps $g_i(\underline{X})$; $i, j = 1, 2, \ldots, R$ --$i \neq j$.

2.4.2 Linear Discriminants

A class of discriminant functions with a particularly simple physical implementation is the linear discriminant function:

$$\begin{aligned} g_i(X) &= \sum_{j=1}^{d} w_{ij}\, x_j + w_{i,d+1} \quad i = 1, 2, \ldots, R \\ &= \underline{W} \cdot \underline{X} + w_{i,d+1} \end{aligned} \tag{2-11}$$

For the special case of R = 2, the relationship of equation (2-3) applies. Substituting (2-11) into (2-3) yields:

$$\begin{aligned} g(\underline{X}) &= \underline{W}_1 \cdot \underline{X} + w_{1,d+1} - \underline{W}_2 \cdot \underline{X} - w_{2,d+1} \\ &= \underline{W} \cdot \underline{X} + w_{d+1} \end{aligned} \tag{2-12}$$

It is noted that the relationship $\underline{W} \cdot \underline{X} = 0$ defines a hyperplane in E^d which passes through the origin. $\underline{W} \cdot \underline{X} + w_{d+1} = 0$ defines a hyperplane with the distance to the origin dependent upon $|\underline{W}|$ and w_{d+1}.

It will be convenient at this point to define an augmented vector $\underline{X}' = (x_1, x_2, \ldots, x_d, 1)$. Equation (2-12) then takes the form:

$$g(\underline{X}) = \underline{W} \cdot \underline{X}' \tag{2-13}$$

It is to be noted that, although $\underline{X}$ and $\underline{W}$ are vectors in E^{d+1}, equation (2-13) set equal to zero defines a dichotomizing hyperplane in E^d. A pattern set for which a hyperplane exists such that all points belonging to category 1 are on one side of the hyperplane and all points belonging to category 2 are on the opposite side of the hyperplane is said to be linearly separable. Given such a training set, training consists of finding a $\underline{W}$ which satisfies the following relationship:

$$\begin{aligned} g(\underline{X}) = \underline{W} \cdot \underline{X}' &> 0 \text{ for all } \underline{X} \in \text{class 1} \\ &< 0 \text{ for all } \underline{X} \in \text{class 2} \end{aligned} \tag{2-14}$$

A device capable of physically implementing equation (2-14) is the threshold logic unit (TLU). The TLU has d inputs which are weighted by multiplying each input by a constant. The weighted inputs are then summed and fed into a threshold device. If the summed input is greater than a pre-set bias level, the output of the threshold device is a high level; if the summed weighted inputs are below the bias level, the output of the threshold device is a low level. For our purposes, the high level

output corresponds to a category 1 membership decision while the low level corresponds to a category 2 membership decision. The inputs are the d measures of the unaugmented pattern and the weights correspond to the first d components of $\underline{W}$. The bias level is associated with w_{d+1}.

An example of a circuit which comprises a TLU is a resistive summing network feeding a high gain amplifier driven into saturation.

Summarizing the properties of a TLU, the dichotomizing surface in E^d is a hyperplane which has an orientation given by the weights w_1, w_2, $\ldots$, w_d with a position proportional to w_{d+1}. The distance from the hyperplane to an arbitrary pattern $\underline{X}$ (unaugmented) is proportional to the value of $g(\underline{X})$.

2.4.3 Piecewise Linear Machines

Another concept which will be of importance in the outlined research is that of a piecewise linear machine. Minimum distance classifiers with respect to point sets constitute a sub-class of piecewise linear machines and will be used for illustrative purposes.

Suppose that it has been determined that the patterns to be classifed cluster (are "close to") about some predetermined set of prototype points. In the discussion which follows $\underline{X}$ will be unagumented. Consider the special case of R categories, each associated with a single point $\underline{P}_i$, i = 1, 2, . . . , R. A reasonable decision rule for classification of given pattern $\underline{X}$ is to assign $\underline{X}$ to that class i_o for which the distance between $\underline{X}$ and the

$\underline{P}_{io}$ is less than the distance between $\underline{X}$ and any other $\underline{P}_i$. Equivalently, one may compare the squared distances.

The squared Euclidian distance between points $\underline{X}$ and $\underline{P}_i$ is given by:

$$(\underline{X} - \underline{P}_i)^2 = (\underline{X} - \underline{P}_i) \cdot (\underline{X} - \underline{P}_i) \qquad (2\text{-}15)$$

$$= \underline{X} \cdot \underline{X} - 2\underline{X} \cdot \underline{P} + \underline{P}_i \cdot \underline{P}_i \qquad (2\text{-}16)$$

It is seen that a choice of the minimum squared distance is equivalent to the selection of the maximum:

$$g_i(\underline{X}) = \underline{X} \cdot \underline{P}_i - 1/2\, \underline{P}_i \cdot \underline{P}_i \qquad (2\text{-}17)$$

Comparing equation (2-17) with equation (2-11), it is seen that:

$$\begin{aligned} w_{ij} &= p_{ij} \qquad & i &= 1, 2, \ldots, R \\ & & i &= 1, 2, \ldots, d \\ w_{i,d+1} &= -1/2\, \underline{P}_i \cdot \underline{P}_i \end{aligned} \qquad (2\text{-}18)$$

and the decision criterion is as previously outlined.

Suppose that, rather than a single point typifying a class, L_i prototype points are associated with class i, $i = 1, 2, \ldots, R$. We may define the R discriminant functions:

$$\begin{aligned} g_i(\underline{X}) &= \max\, [g_i^{(j)}(\underline{X})] \\ &= \max\, [\underline{P}_i^{(j)} \cdot \underline{X} - 1/2\, \underline{P}_i^{(j)} \cdot \underline{P}_i^{(j)}] \qquad (2\text{-}19) \\ & \qquad j = 1, 2, \ldots, L_i \\ & \qquad i = 1, 2, \ldots, R \end{aligned}$$

The $g_i^{(j)}(\underline{X})$ is called a subsidiary discriminant function and is seen to be of the form of equation (2-17) which is linear in $\underline{X}$. Since each of the $g_i(\underline{X})$ is a piecewise linear function of $\underline{X}$ they are called piecewise linear discriminant functions.

The decision surfaces which they describe are sections of hyperplanes. It is noted that for the particular constraints imposed that the decision regions described are convex.

2.4.4 Layered Machines

It has been shown[10] that for the case of $R = 2$, a special class of circuits called layered machines implement a piecewise linear machine.

A layered machine is a network of TLUs organized in layers so that the inputs to each bank of TLUs are the outputs of the preceeding bank, with the exception of the first bank, whose inputs are the pattern to be classified. The last bank consists of a single TLU.

2.5 General Model Summary

This chapter of this paper has reviewed those topics in pattern recognition theory which have a direct relation to the particular pattern classifier which was implemented. The specific pattern classifier used in the research is described in Chapter 5.

CHAPTER III

DATA ACQUISITION AND FEATURE CALCULATION

3.1 Introduction

A decision as to what to measure must be made early in an applied pattern recognition program. The type of feature utilized will, of necessity, vary depending upon the particular application. For the case of cough categorization, the physiological model of the cough reflex would be expected to indicate candidate features. This physiological model is presented in this chapter of this report and is followed by a description of data acquisition methods and an outline of the algorithms used for feature calculation.

3.1 Physiological Considerations

3.1.1 The Respiratory Mechanism

Air enters the respiratory system via the oral or nasal cavities which open into the pharynx. The pharynx separates into the trachea and the esophagus directly above the larynx. At the point of division, food is separated from air, food being diverted to the esophagus by the closure of the epiglottis, a flap which closes over the opening of the trachea when food touches the pharynx. When the epiglottis is not blocking the opening to the trachea the lower respiratory tract presents less resistance to

the flow of air than the path through the esophagus. Ventilation is thus provided to the lungs.

The larynx is located at the top of the trachea. The vocal cords are the portion of the larynx that produce sound. These cords are two small vanes located on either side of the air passageway. The vanes meet at an angle at the back of the larynx. As the muscles of the larynx are contracted, closure of the vanes starts at the point of intersection and works its way to the base of the triangle they form. During phonation they essentially close off the trachea. Air forced through the closed vanes sets up a lateral vibration which modulates the air stream at audible frequencies. The frequency of vibration is determined by the degree of contraction of muscles in the larynx.

The trachea continues below the larynx to a point beneath the top of the lungs where it branches into the bronchial system. The bronchi are tubes of varying diameter, each bronchus branching into a network of smaller bronchi, thereby forming the "bronchial tree". The smallest of the bronchi branch into the alveolar ducts which connect to the alveolar sacs via atria. The alveolar sacs are encased in the pulmonary membrane, which has a thickness of from 1 to 4 microns (several times less than the thickness of a red blood cell). It is in the alveolar sacs that the oxygen-carbon dioxide exchange takes place between the blood and the air in the atria.

The lungs are enclosed in the thoracic cage. A negative pressure exists in the pleural cavity (a potential space between the lungs and chest wall) with respect to the interior of the lungs. An expansion of the thoracic cage therefore expands the lungs, while a contraction of the cage produces expiration.

The major muscles which enter into inspiration are the diaphragm, the external intercostals, and a number of small muscles in the neck. The downward movement of the diaphragm pulls the bottom of the pleural cavity downward (thereby elongating it) while the external intercostals and neck muscles lift the front of the cage, causing the ribs to angulate forward (increasing the thickness of the cage).

The major muscles of expiration are the abdominals and, to a lesser extent, the internal intercostals. The abdominal muscles pull downward on the chest cage (decreasing the thoracic thickness) and force the abdominal contents upward against the diaphragm (decreasing the longitudinal dimension of the pleural cavity). The internal intercostals aid slightly in expiration by pulling the ribs downward (decreasing the thickness of the chest).

3.1.2 The Cough Reflex

The cough reflex provides a means for the body to clear its airways. It is triggered by an irritant touching the surface of the glottis, the trachea, or a bronchus. The respiratory muscles first contract very strongly building up high pressure in the lungs while simultaneously, the epiglottis blocks

the trachea and the vocal cords clamp tightly closed. These impediments to air flow are suddenly removed. This allows the pressurized air to flow out of the lungs at high velocities carrying unwanted particles with it.

Air flow and pressure measurements have been made during coughs on normal subjects and patients suffering from various respiratory diseases. A study by Whittenberger and Mead[14] indicated that peak interthoracic pressure relative to atmospheric pressure ranged between 90 and 152 mm hg on the six subjects used (three normal, one asthmatic, two with emphysema of long standing). Maximum flow rates ranged between 480 and 700 liters per minute for the normals and 25 to 150 liters per minute for the patients with emphysema. The total volume expired during the first 0.2 sec of the cough ranged from 0.15 liter for one of the patients with emphysema to 1.19 liters for one of the normals. The pressure was sustained for a considerably longer period of time for the diseased subjects than for the normal, indicating a higher impedance to flow and decreased efficiency of the cough.

Other studies have radiologically examined the outline of the trachea and bronchi during cough. There seems to be no question as to whether or not these passages contract during the cough, but precisely what mechanism is involved is in doubt. It has been shown experimentally that the airway passages contract even during normal expiration.

DiRienzo[7] postulated a true peristaltic action which aided in the expulsion of the irritant. Other researchers disagree, stating that the contraction, although variable along the length of the airways, is not a coordinated peristaltic action at all, but suggest that the contraction is simply a means for increasing the laminar air velocity to that required for efficient cleansing of the airways.

There is some indication that the increased resistance to flow is highly localized. Persons suffering from emphysema, asthma, and bronchitis are unable to achieve these high velocities. A study was performed in which a bronchiodilator was administered to patients suffering from bronchitis, others having emphysema, and a group of normals for control.[1] It was found that the air velocity increased for those with bronchitis, was unchanged for those with emphysema, and was unchanged for the normals. This gives some indication as to the general origin of the resistance to flow in the two disease classes.

3.2 Data Acquisition

3.2.1 Recording

The data with which this paper is concerned were obtained in the form of analog tape recordings. The recordings used fall into two categories -- those recorded directly from hospital rooms at Woodlawn Hospital in Dallas, Texas and recordings of forced coughs and simulated artifacts.

The hospital recordings were obtained with the following equipment arrangement. Microphones were placed in selected hospital rooms, primarily in closets. The signals from the microphones were carried, via long cables, to recording apparatus in the basement of the building. The signals from the microphones were recorded on a continuous tape loop. The amplified signal from the microphones were operated on by a voice controlled relay connected in parallel with the continuously operated tape recorder. A filter was adjusted empiricially so that the voice controlled relay threshold was not exceeded by all artifacts, but it was not possible to efficiently separate all coughs and artifacts by this adjustment. The voice controlled relay actuated a second tape recorder which was operated in a start--stop mode. The continuous loop recorder output provided the input for the start--stop recorder and allowed sufficient delay for the second recorder to come up to speed after having been actuated by the voice controlled relay. Recording speed of that start--stop recorder was 3 3/4 ips.

Recordings were made by the above described equipment mostly at night, which is a period of light hospital activity, to avoid recording the numerous acoustical artifacts which occur during the normal daylight routine. The equipment was unattended for the majority of the time that recording took place.

The recordings so obtained are typically characterized by high background noise (with respect to acoustical events occurring in the hospital rooms) and widely varying recording levels. Fidelity was not at the most desirable level because of the low recording speeds and the re-recording. For the most part, however, coughs and artifacts could be distinguished by ear.

Recordings of simulated coughs and artifacts were made both in the laboratories at Woodlawn Hospital and at The University of Texas at Austin. Recording speed was 7 1/2 ips and recording levels were monitored. Coughs were forced coughs of normal subjects. Some artifacts, such as conversation and doors closing, were recorded directly. Others were obtained from commercial sound-effects recordings.

3.2.2 Digitization and Segmentation

Early in the research program it was necessary to make a decision as to the basic philosophy to be followed in data reduction. Two alternatives were apparent. The first was to build special purpose circuitry to pre-process the recorded audio signals; the second was to digitize the audio data directly and perform all data reduction on a general purpose computer. The second alternative was chosen in the interests of flexibility.

Two computer facilities were utilized. The Electrical Engineering Department at the University maintains a Scientific Data Systems 930

computer which includes as peripheral equipment analog-to-digital and digital-to-analog converters. The University computation center maintains a Control Data Corporation 6600 computer. The SDS 930 is operated on an open shop basis while the CDC 6600 computer is operated on a closed shop basis. Data digitization was performed by the SDS 930 computer while the majority of the computations were accomplished by the CDC 6600 computer.

Prior to digitization of the data, spectrograms were made of a representative sample of the analog signals. It was found that the highest frequency of significant magnitude was in the vicinity of 6 Khz. The Nyquist rate is therefore approximately 12 K samples/sec. In practice it would be desirable to have a sampling rate somewhat in excess of this value. A program was written for the SDS 930 computer which allowed digitization at a sampling rate of 16,500 samples/sec.

The A/D converter performs a 12 bit conversion with a maximum conversion rate in excess of 30 K conversions/sec. Conversion accuracy is specified by the manufacturer as $\pm$ the least significant bit. Full scale input used was $\pm$ 10v to yield a quantization error of approximately 10 mv. The central processor memory consists of 8,192--24 bit words. The nominal length of a signal to be digitized is two seconds. At the 16,500 conversions/sec rate used, the available memory is not adequate to buffer the digitized data.

The magnetic tape units which are included in the peripheral equipment may be operated in an interlaced mode of operation, however. When operated in this fashion access to the central processor is required only when data is required from the memory. The interlaced mode utilizes a priority interrupt system. One of the priority interrupts is available for program use in a real-time environment. This interrupt was used to signal the beginning of a conversion.

The A/D conversion program controls time of sampling to within - 0, + 3.5 microseconds. The digital data is formatted and buffered in the central processor memory while previously converted data is simultaneously written on digital magnetic tape. The conversion rate is limited by the speed of the tape transport. Although a higher conversion rate could be accomplished at an 800 bpi data density, 556 bpi recording density was used in the interests of greater reliability.

A digital-to-analog conversion sub-routine was included in the program. Use of the interlace feature was made to allow simultaneous D/A conversion and reading of the digital tape. It was found that the D/A output could be directly connected to a speaker to yield an audible output with sufficient fidelity for monitoring purposes. Additional D/A features were incorporated to output synchronization and calibration signals for oscilloscope or oscillograph monitoring.

The tape recorder output was connected to the input of an adjustable filter with attenuation characteristics of 24 db per octave. The break frequency was set at 4 kHz. The output of the filter was connected to the computer multiplexer -- A/D converter input.

As previously stated, conversions were initiated by a priority interrupt. A pulse generator provided the input signal to the interrupt. The pulserepetition rate was 16,500 pps, ± 1 pps. The output of the pulse generator was continously monitored by a digital counter.

Conversion was initiated and ended by computer console sense switch operation. After an A/D conversion was completed confirmation of the digitized signal was made by utilization of the D/A conversion subroutine.

The digital data tape which was the product of the program was formatted with 1,000 -- 12 bit conversions per record. The binary representation was integer two's complement. An identification record followed data records within a file, a file consisting of the data converted during an A/D conversion cycle. Adjacent files were separated by End of File marks with the last file on the tape being followed by double End of File marks.

The CDC 6600 computer's binary representation of integer data is in one's complement, sixty bits per word. The 6600 was programmed to reformat the data.

Varying recording levels in the original analog data required that the data be normalized with respect to amplitude. This was accomplished by arbitrarily setting the datum point in a file with the greatest magnitude to a floating point magnitude of 10.0 and scaling the remainder of the data points in the file accordingly. Additionally, efficient computation by subsequent programs required that the files be reformatted so that the identification record preceeded the data in the file. Provision was also made to ensure that the mean of the data was 0. The normalized data was written in binary format on magnetic tape.

Due to the manual start - - stop digitizing technique used and the nature of the original analog data, files contained segements of low level data (backgound noise). It was therefore necessary to locate segments of the files which contained usable data. An algorithm subsequently used to calculate the Fourier series representation of the time varying signal required that segments of data used in the calculation contain a number of data points exactly divisible by a power of two (this algorithm is described in a later section of the dissertation). The divisor chosen was 1024 points. A file was searched until a value exceeding a specified threshold L, was encountered. The conversion number of that value was assigned as the start of a segment. Subsequent data points were assigned to the segment until no values exceeding L were found for a pre-specified time interval, t_d. The number of conversions

corresponding to t_d were subtracted from the total number of points assigned to the segment to obtain the length of the segment. If the segment so located was shorter than a specified time limit, t_r, the segment was rejected. After segmentation of a file was completed, segment lengths were adjusted to be an integral multiple of 1024 points. The values L, t_r, and t_d were determined empirically. The following values were found to give satisfactory results:

$$L = 2.0$$

$$t_d = t_r = 1024/16500 \text{ sec} \approx 1/16 \text{ sec}$$

The normalizing factor for eac.. . was recorded for future use in calculation of the significant noise level.

After segmentation was accomplished on the CDC 6600, the original data was scanned on the SDS 930 by the D/A conversion sub-rountine. Class identification of the segments and 1024 point subsegments was confirmed by listening to the D/A converted data.

Summarizing the digitization and segmentation programs, analog data was digitized on the SDS 930 at a conversion rate of 16500 conversions/sec. The data obtained was normalized with respect to amplitude and subsegmented on the CDC 6600 computer. Confirmation of segmentation and class identification (cough or artifact) was then made by use of the SDS 930 D/A conversion subroutine.

After the completion of the above process the digitized data was on magnetic tape in a format compatible with computation in the CDC 6600. Punched cards were included in the output from the 6600 program on which were entered segment and class (cough or artifact) identification. After the original digitized data was scanned on the SDS 930 to confirm class identification of the subsegments, the punched cards were altered if original class identification was erroneous (a section of background noise identified as a cough). The punched cards were then used as input data, along with the 6600 formatted tape, for subsequent programs on the CDC 6600. Data not included within one of the subsegments was not considered in subsequent calculations.

3.3 Feature Extraction

The measures calculated are described in sections 3.3.1 through 3.3.4 of this report. They fall into four broad categories:

1) those concerned with the magnitude of the normalized amplitude of the signal.
2) those which represent the form of the envelope of the signal.
3) zero crossing representations.
4) spectral analysis measures.

3.3.1 Amplitude Density Approximation Feature

An examination of the waveform of the recorded signals indicated that an analysis of the shape of the amplitude distribution of the normalized

signals could offer possibilities as a measure.

A 1024 point subsegement of data was operated on to obtain an est.mate of the probability density function of the amplitude of the signal. Thirty-one equi-spaced limits were used to yield thirty-two increments. The increment limits are as shown in Table 3-1. It is noted that normalization accomplished during reformatting was with respect to the point with maximum magnitude within a recorded file and that several segments and subsegments were included within this file. The mean value of all points within a file was zero, but such was not necessarily the case within a subsegment.

The algorithm used to implement the amplitude density approximation is given below. The increment address calculations take the form of a tree where the iterative decisions described determine the branch chosen. The tree terminates with a total number of branches equal to the number of classification intevals.

Define:

$I_j \triangleq$ value of jth sample in a subsegment

Number of intervals used $\triangleq 2^{M+1}$

$D \triangleq$ interval width (difference of interval limits)

$V \triangleq 2^M D$

$L \triangleq$ interval number; highest amplitude value, lowest number; $L = 1, 2, \ldots, 2^{M+1}$

Table 3-1

Increment Limits

Amplitude Density and Envelope Amplitude Density

Features

Increment No.	Lower Limit	Upper Limit	Increment No.	Lower Limit	Upper Limit
1		-9.375	17	0.000	0.625
2	-9.375	-8.750	18	0.625	1.250
3	-8.750	-8.125	19	1.250	1.875
4	-8.125	-7.500	20	1.875	2.500
5	-7.500	-6.875	21	2.500	3.125
6	-6.875	-6.250	22	3.125	3.750
7	-6.250	-5.625	23	3.750	4.375
8	-5.625	-5.000	24	4.375	5.000
9	-5.000	-4.375	25	5.000	5.625
10	-4.375	-3.750	26	5.625	6.250
11	-3.750	-3.125	27	6.250	6.875
12	-3.125	-2.500	28	6.875	7.500
13	-2.500	-1.875	29	7.500	8.125
14	-1.875	-1.250	30	8.125	8.750
15	-1.250	-0.625	31	8.750	9.375
16	-0.625	0.000	32	9.375	

$X_j \triangleq$ dummy variable used in calculation of interval address of the jth sample point

The following iterative algorithm identifies the interval address of a sample with a total of M + 1 tests. Parenthesized superscripts denote iteration number:

$$X_j^{(0)} \triangleq \Gamma_j \quad ; \quad L_j^{(0)} \triangleq 1$$

If:

$$X_j^{(i-1)} - V/2^i < 0 \; ; \quad L_j^{(i)} = L_j^{(i-1)} + 2^{M-i}$$

$$X_j^{(i)} = X_j^{(i-1)}$$

$$\geq 0 \; ; \quad L_j^{(i)} = L_j^{(i-1)}$$

$$X_j^{(i)} = X_j^{(i-i)} - V/2^i$$

$$i = 1, 2, \ldots, M$$

$$\Gamma_j < 0 \quad ; \quad L_j^{(M+1)} = 2^{M+1} - L_j^{(M)}$$

$$\geq 0 \quad ; \quad L_j^{(M+1)} = L_j^{(M)}$$

It is noted that this algorithm is most efficient if the Γ_j's are uniformly distributed between $\pm$ V.

The CDC 6600 is capable of calculating several arthmetic operations simultaneously providing that the calculations are independent. When a calculation is dependent upon the outcome of a test, such is not the case. Minimization of the number of tests is therefore roughly equivalent to

minimization of calculation time. Table lookups were employed to obtain 2^i and $V/2^i$ in the above algorithm so that these time consuming calculations were not made during every iteration.

After the above algorithm operated on all data points in a subsegment, the frequency of occurrence values were normalized in order to estimate the amplitude density functions by dividing each interval count by the total number of points classified. For convenience in later calculations, the array storing the normalized values was then inverted so that the lowest numbered interval corresponded to the most negative limit.

3.3.2 Envelope Descriptive Features

A preliminary study of the waveform of the recorded coughs and artifacts indicated that the set of candidate features should include envelope descriptive measures.

The cough signal started with a large amplitude and died off quasi-exponentially, as would be expected from the physiological model. This was not the case for the majority of the waveforms of recorded artifacts. The two envelope dependent features which were calculated were an estimation of the probability density of the envelope amplitude and an estimation of the probability density of the slope of the envelope.

3.3.2.1 Envelope Amplitude Density Feature

In order to estimate the probability density function of the amplitude of the envelope, a sampled representation of the envelope was required.

Initially the peak values of the signal, which were points on the envelope, were obtained by applying Sterling's interpolation approximation from numerical analysis. The time of maxima or minima were computed by differentiation of Sterling's formula and then solving for the time when the derivative of the approximation is zero. A re-application of Sterling's formulation yielded the interpolated value of the amplitude of the signal at the time of occurrence of the extrema.

Sterling's approximation attempts to fit a polynomial to the discrete points in the neighborhood of the point of interest. Having evaluated the coefficients of the polynomial, one solves for the value at the point of interest. A comparison of the values obtained by this approach to that obtained by noting the time of occurrence and the amplitude of the peak sampled points indicated that little difference existed. In the interests of computational efficiency, the algorithm used for extraction of a sampled representation of the envelope used this latter approach.

A 1024 point subsegment was scanned to locate the maximum and minimum points. The time of occurrence and amplitude of the extreme were stored in central memory. The algorithm utilized to obtain the peak points and times of occurrence is described below. It will be noted that the maxima are positive and the minima negative.

Define:

$T \triangleq$ magnitude of maximum peak to peak noise

$F_j \triangleq$ value of jth data point (relative to start of subsegment)

$N_p \triangleq$ number of positive maxima

$N_n \triangleq$ number of negative minima

$D \triangleq$ dummy variable

$E_{p\ell} \triangleq$ amplitude of ℓ^{th} positive maximum

$E_{nm} \triangleq$ amplitude of n^{th} negative minimum

$M_{p\ell} \triangleq$ sample number of ℓ^{th} positive maximum relative to start of subsegment (proportional to time of occurrence)

$M_{nm} \triangleq$ sample number of m^{th} minimum relative to start of subsegment

The following algorithm identifies and stores the quantities $E_{p\ell}$, E_{nm}, $M_{p\ell}$ and M_{nm}. The quantities N_p and N_n are computed. Parenthesized superscripts denote iteration number:

$$D^{(o)} \triangleq \ , \ \ell^{(o)} \triangleq 0, \ m^{(o)} = 0, \ N_p^{(o)} = 0, \ N_n^{(o)} = 0$$

Define the following events:

$$A^{(i)} \triangleq \left\{ \xi : D^{(i-1)} \neq 0, \ |F_{i+1} - F_i| > T, \ \mathrm{Sgn}\,(F_{i+1} - F_i) \neq \mathrm{Sgn}\, D^{(i-1)} \right\}$$

$$G^{(i)} \triangleq \left\{ \xi : (F_{i+1} - F_i) > 0 \right\}$$

$$H^{(i)} \triangleq \left\{ \xi : F_i < 0 \right\}$$

$$B^{(i)} \triangleq G^{(i)} \cap H^{(i)}$$

$$C^{(i)} \triangleq \overline{G}^{(i)} \cap \overline{H}^{(i)}$$

Define the indicator functions:

$$I_A^{(i)} = \begin{cases} 0 & \text{if } \xi \not\in A^{(i)} \\ 1 & \text{if } \xi \in A^{(i)} \end{cases}$$

$$I_B^{(i)} = \begin{cases} 0 & \text{if } \xi \not\in B^{(i)} \\ 1 & \text{if } \xi \in B^{(i)} \end{cases}$$

$$I_C^{(i)} = \begin{cases} 0 & \text{if } \xi \not\in C^{(i)} \\ 1 & \text{if } \xi \in C^{(i)} \end{cases}$$

If $I_A^{(i)} I_B^{(i)} = 1$ then $m^{(i)} = m^{(i+1)}$

$$M_{nm}^{(i)} = i$$

$$E_{nm}^{(i)} = F_i$$

$$D^{(i)} = +1$$

$$N_n^{(i)} = N_n^{(i-1)} + 1$$

If $I_A^{(i)} I_C^{(i)} = 1$ then $\ell^{(i)} = \ell^{(i-1)} + 1$

$$M_{p\ell}^{(i)} = i$$

$$E_{p\ell}^{(i)} = F_i$$

$$D^{(i)} = -1$$

$$N_p^{(i)} = N_p^{(i-1)} + 1$$

Otherwise quantities are unchanged from the previous iteration.

After the peak points have been obtained, the E_p's and E_n's are used to estimate the probability density of the envelope amplitude by implementation of the algorithm described in Section 3.3.1. Thirty-two increments were utilized in the approximation. Limit values were as given in Table 3-1.

3.3.2.2 Envelope Difference Density Feature

The slope of the envelope was approximated by dividing the difference in amplitude of adjacent extreme points by the number of sampling intervals between these points. All quantities required for these calculations were obtained by use of the algorithm described in Section 3.3.2.1.

The maximum and minimum points were operated on separately. Prior to calculation of the envelope differences for the negative envelope, these minima were replaced by their negatives (to yield all positive points).

After calculation of the differences,an estimate of their probability density was made by application of the algorithm outlined in Section 3.3.1. Sixty-four intervals were used in the approximation with limits as shown in Table 3-2.

3.3.3 Zero-Crossing Features

A measure which is frequently used because of its computational simplicity is a count of zero crossings of a signal of a signal during a specified time interval. Alternatively,measurement of the periods between zero crossings would contain more complete information.

The algorithm that was implemented calculated the time interval (to the nearest sample period) between zero crossings as a first step. A dead band (centered at zero) which exceeded the maximum noise level was assigned prior to implementation of the algorithm. It was required that the signal pass through the dead band before a zero crossing was

Table 3-2

Increment Limits

Envelope Difference

Feature

Increment No.	Upper Limit	Increment No.	Upper Limit	Increment No.	Upper Limit	Increment No.	Upper Limit
1	-1.695	17	-0.820	33	0.055	49	0.930
2	-1.641	18	-0.766	34	0.109	50	0.984
3	-1.586	19	-0.711	35	0.164	51	1.039
4	-1.531	20	-0.656	36	0.219	52	1.094
5	-1.477	21	-0.602	37	0.273	53	1.148
6	-1.422	22	-0.547	38	0.328	54	1.203
7	-1.367	23	-0.493	39	0.383	55	1.258
8	-1.313	24	-0.438	40	0.438	56	1.313
9	-1.258	25	-0.383	41	0.493	57	1.367
10	-1.203	26	-0.328	42	0.547	58	1.422
11	-1.148	27	-0.273	43	0.602	59	1.477
12	-1.094	28	-0.219	44	0.656	60	1.531
13	-1.039	29	-0.164	45	0.711	61	1.586
14	-0.984	30	-0.109	46	0.766	62	1.641
15	-0.930	31	-0.055	47	0.820	63	1.695
16	-0.875	32	0.000	48	0.875		

tabulated.

Memory locations corresponding to logarithmically spaced bands, limits of which are given in Table 3-3, were incremented in accordance with the number of sampling increments between zero crossings. After a 1024 point subsegment of data had been operated on, the values for the zero crossing period counts were normalized by dividing the number of tabulated crossings per interval by the total number of zero crossings for the 1024 point subsegment to yield an estimate of the probability density of the zero crossing periods. Additionally, the average zero crossing frequency was obtained from the total number of zero crossings tabulated.

3.3.4 Spectral Features

As previously noted, spectrograms of selected recorded audio signals were made and examined for characteristics which were unique to the audible cough as contrasted with the artifacts. It was expected that resonances caused by the various cavities in the vocal tract could possibly give rise to broad frequency peaks analogous to formants encountered in speech recognition studies.[8] Formant like structures were not apparent, however.

Although formants were not present, information would still be expected to be contained in a frequency domain analysis of the signal. Several approaches to obtaining this analysis were apparent. A Fourier

Table 3-3

Band Limits

Zero Crossing Period

Feature

Band No.	Lower Frequency* Limit	Upper Frequency* Limit
1	0	16.11
2	16.11	32.23
3	32.23	64.45
4	64.45	1.289×10^2
5	1.289×10^2	2.578×10^2
6	2.578×10^2	5.156×10^2
7	5.156×10^2	1.031×10^3
8	1.031×10^3	2.063×10^3
9	2.063×10^3	4.125×10^3
10	4.125×10^3	

*Frequency = 1/period Hz

series representation or power spectral density estimate over a sub-segment would be candidates. An alternative method of obtaining similar information would be to [illegible] a digital band-pass filtering on the time varying signal directly and to then obtain the power output from a digital band-pass filter over segments of time.

All of the above computations require a large number of multiplications. Techniques have been developed for efficient estimation of the power spectal density[2] and recursive algorithms could be implemented in the case of the digital filters. Computation time would still be great, however.

The advent of a fast Fourier algorithm[4] allowed direct calculation of the complex Fourier series coefficients of subsegments of the data. This algorithm was programmed for the CDC 6600 computer. Since the program follows the published Cooley-Tuckey algorithm step by step, it will not be described.

A 1024 point subsegment of data was operated on to yield 511 (excluding the DC component) usable complex coefficients. The numbers obtained were proportional to, rather than equal to, the Fourier coefficients since divisions necessary to take into account the sampling rate and period of the subsegment were not performed; sampling rates and subsegment lengths

are identical for all data processed and the values are required for comparative purposes only.

Because of the large number of coefficients which resulted from the computations outlined above, it was necessary to effect further data reduction. This was done by summing the squared magnitudes of the coefficients over bands and by extracting information concerning peak frequencies and amplitude of the magnitude squared of the coefficients at these peak frequencies.

3.3.4.1 Fourier Band Features

As noted above, the squared magnitudes of the Fourier coefficients were summed over bands. A total of nine logarithmically spaced bands were used. Band limits are shown in Table 3.4.

3.3.4.2 Frequency Peak Features

The squared magnitude of the Fourier coefficients were operated on by the envelope determining algorithm described in Section 3.3.2.1 of this report to determine the frequencies at which amplitude peaks occurred. These maxima were then ordered with respect to magnitude and the twenty peaks with the greatest magnitude were selected. A total of eighty measures were extracted in the form of the magnitude of the twenty highest peaks, the twenty frequencies corresponding to these peaks and the frequencies on either side of the peak frequencies at which the squared magnitude was one-fourth of the peak value. In the

Table 3-4

Fourier Coefficient

Magnitude Squared Bands

Frequency Limits

Band No.	Upper Limit (Hz)
1	16.11
2	48.34
3	1.128×10^2
4	2.417×10^2
5	4.995×10^2
6	1.015×10^3
7	2.046×10^3
8	4.109×10^3
9	8.234×10^3

event that less than twenty peaks were apparent, a value of zero was assigned to the measures for which no peaks were found.

3.3.5 Measure Summary

Table 3-5 lists the features which were calculated and the measure numbers assigned to them.

3.4 Measure Selection

The algorithms outlined in Section 3.3 of this paper were used to calculate a total of 228 measures. A segment of data contains several subsegments, each of which is represented by these 228 quantities. If the segment consisted of thirty subsegments (a not unusual occurrence), the pattern would have 6840 dimensions. Such a large dimensionality is unfeasible for real time implenentation by special purpose circuitry. The fast access memory requirements for a general purpose computer implementation would be sizable. It was therefore necessary to select a subset of the computed measures for use in training and classification.

The problem, then, is to order the measures in a manner which reflects the maximum probability of correct classification or, equivalently, which minimizes the probability of wrong classification. After ordering the measures one would choose the number of ordered measures which either allows classification with a tolerable error or which fills the maximum allowable capacity of the machine.

Table 3-5

Table of Measures

Measure Nos.	Feature
1--32	Amplitude density -- least values, lowest number
33--64	Envelope amplitude density -- least values, lowest number
65-128	Envelope difference density -- least value, lowest number
129-138	Zero crossing period density -- lowest frequency band, lowest number
139	Average zero crossing frequency
140-148	Magnitude squared Fourier coefficient bands, lowest frequency, lowest number
149-168	Normalized frequency of Fourier peaks -- highest amplitude peak, lowest number
169-188	Amplitude of Fourier peaks corresponding to measures 149--168
189-208	Lower quarter power frequency corresponding to measures 149-168
209-228	Upper quarter power frequency corresponding to measures 149-168

If the pattern is the input to a machine which depends upon distance in some sense between patterns of differing classes to effect recognition, one wishes to identify those measures for which high pattern densities in E^d of the opposing classes are least overlapping. In terms of the joint probability density functions this indicates that the chosen measures should be those for which the joint probability density functions, conditioned on class membership, exhibit the greatest difference in value.

If a Bayes machine is to be implemented, it can be shown that the same criterion applies. For the two class case, recall that the classification surface of the Bayes machine with a symmetric loss function is given by:

$$p(1)p(\underline{X}|1) - p(2)p(\underline{X}|2) = 0 \tag{3-1}$$

Define:

$$R_1 \triangleq \{\underline{X}: p(1)p(\underline{X}|1) > p(2)p(\underline{X}|2)\}$$

$$R_2 \triangleq \{\underline{X}: p(2)p(\underline{X}|2) > p(1)p(\underline{X}|1)\}$$

Assuming continuous probability density functions, the probability of making an erroneous decision, P_e, is given by equation (3-2).

$$\begin{aligned} P_e = & \int \cdots \int_{R_1} p(2)p(\underline{X}|2)\, dx_1 \ldots dx_d \\ & + \int \cdots \int_{R_2} p(1)p(\underline{X}|2)\, dx_1 \ldots dx_d \\ = & \int \cdots \int_{R_1} p(1)p(\underline{X}|1)\, dx_1 \ldots dx_d \end{aligned} \tag{3-2}$$

$$- \int \cdots \int_{R_1} \left[p(1)p(\underline{X}|1) - p(2)p(\underline{X}|2) \right] dx_1 \ldots dx_d$$

$$+ \int \cdots \int_{R_2} p(2)p(\underline{X}|2) \quad dx_1 \ldots dx_d$$

$$- \int \cdots \int_{R_2} \left[p(2)p(\underline{X}|2) - p(1)p(\underline{X}|2) \right] dx_1 \ldots dx_d$$

$$P_e = \int \cdots \int_{R_1} p(1)p(\underline{X}|1) \, dx_1 \ldots dx_d$$

$$+ \int \cdots \int_{R_2} p(2)p(\underline{X}|2) \, dx_1 \ldots dx_d \qquad (3\text{-}3)$$

$$- \underbrace{\int_{\infty}^{\infty} \cdots \int_{\infty}^{\infty}}_{\substack{\text{All} \\ E^d}} \left| p(1)p(\underline{X}|1) - p(2)p(\underline{X}|2) \right| dx_1 \ldots dx_d$$

Minimization of P_e implies maximization of α, defined in equation (3-4) given below:

$$\alpha \triangleq \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{\substack{\text{All} \\ E^d}} \left| p(1)p(\underline{X}|1) - p(2)p(\underline{X}|2) \right| dx_1 \ldots dx_d \qquad (3\text{-}4)$$

If the probability density functions are discrete rather than continuous, summations take the place of integrations.

Use of (3-4) requires knowledge of the a priori probabilities of occurrence and the conditional joint density functions for sub-sets of the measures taken d at a time. Should the a priori probabilities be unknown and unestimable, maximum likelihood classification is appropriate, in

which case the quantity to be maximized is given by β, defined in equation (3-5) given below:

$$\beta \triangleq \underset{\substack{\text{All} \\ E^d}}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}} \left| p(\underline{X}|1) - p(\underline{X}|2) \right| dx_1 \ldots dx_d \tag{3-5}$$

It is seen that β assumes a maximum value of 2 for disjoint classes. Under the presumption of statistical indpendence of the measures it has been shown[3] that minimization of equation (3-5) is equivalent to selection of the d measures which maximize γ, defined as:

$$\gamma \triangleq \sum_{i=1}^{d} \int_{-\infty}^{\infty} \left| p(x_i|1) - p(x_i|2) \right| dx_i \tag{3-6}$$

where $p(x_i \mid 1)$ and $p(x_i \mid 2)$ are the marginal conditional densities. Some measures with which this dissertation is concerned undoubtedly exhibit statistical dependence. However, due to the impracticability of estimating the joint probability functions and uncertainty regarding a priori probabilities of occurrence, the relationship given in equation (3-6) was used to select the sub-set of measures upon which training and recognition was performed.

The algorithm described in Section 3.3.1 of this paper was applied to the measures to estimate their marginal conditional probability density functions. The absolute values of the differences in estimated probabilities for the two classes were then summed and the

resulting values ordered with respect to magnitude. Results are given in Chapter 5.

CHAPTER IV

PATTERN CLASSIFIER MODEL

4.1 General Considerations

Prior to describing the machine which was chosen to implement the categorization, a brief review of the desirable properties of the machine is in order.

Since we prefer that the model chosen be adaptable to real time implementation, simplicity from a hardware viewpoint is a consideration. The TLU previously described would meet this requirement, as would circuitry which selects the maximum of several signals which are weighted sums of the input measures. Networks which realize a logical switching circuit would also be candidates. Other possible implementations would require the storage of values in a memory of some description. While these techniques could be implemented with relative ease in a general purpose computer program, the real time circuitry would be more complex than would be desirable.

It has previously been noted that the use of the general Bayes machine requires estimation (and subsequent storage) of $p(\underline{X}|i)$ and $p(i)$ for all $\underline{X}$ in E^d, which is a formidable task. If, however, the measures to be used are binary valued and statistically independent, the Bayes machine

for the symmetric loss function assumes a particularly simple form. The derivation of the discriminant function for this special case is taken from Nillson[10].

Recall that the Bayes discriminant function for the symmetric loss function is $g_i(\underline{X}) = p(\underline{X}/i)p(i)$, or equivalently with respect to the decisions made, $g_i(\underline{X}) = \ln p(\underline{X}/i) + \ln p(i)$, $i = 1, \ldots, R$. For the case of $R = 2$ and the x_j statistically independent, we may write:

$$g(\underline{X}) = \sum_{i=1}^{d} \ln p(x_i/1) + \ln p(1) - \sum_{i=1}^{d} \ln p\,(x_i/2) - \ln p(2)$$

$$= \sum_{i=1}^{d} \ln \frac{p(x_i/1)}{p(x_i/2)} + \ln \frac{p(1)}{p(2)} \qquad (4\text{-}1)$$

where the x_i may take on values of only 0 and 1.

The quantities which must be estimated have been reduced to the following:

$$\begin{aligned} p(x_i = 1/1) &\triangleq p_i \\ p(x_i = 0/1) &\triangleq 1 - p_i \\ p(x_i = 1/2) &\triangleq q_i \\ p(x_i = 0/2) &\triangleq 1 - q_i \qquad i = 1, \ldots, d \end{aligned} \qquad (4\text{-}2)$$

$$p(1)$$

$$p(2) = 1 - p(1)$$

Making use of the fact that the x_i may take on values of only 0 and 1, we may write:

$$\ln \frac{p(x_i/1)}{p(x_i/2)} = x_i \ln \frac{p_i}{q_i} + (1 - x_i) \ln \frac{1 - p_i}{1 - q_i} \tag{4-3}$$

Substitution of (4-3) into (4-2) yields:

$$g(\underline{X}) = \sum_{i=1}^{d} x_i \ln \frac{p_i(1 - q_i)}{q_i(1 - p_i)} + \sum_{i=1}^{d} \ln \frac{1 - p_i}{1 - q_i} + \ln \frac{p(1)}{1 - p(1)} \tag{4-4}$$

Equation (4-4) is of the form $g(\underline{X}) = \underline{W} \cdot \underline{X}'$ with:

$$w_i = \ln \frac{p_i(1 - q_i)}{q_i(1 - p_i)} \qquad i = 1, \ldots, d$$

$$w_{d+1} = \sum_{i=1}^{d} \ln \frac{1 - p_i}{1 - q_i} + \ln \frac{p(1)}{1 - p(1)} \tag{4-5}$$

Training of the machine would then consist of estimating the values of p_i, q_i, $p(1)$ and $p(2)$ and substituting these estimates into the relationships given in equation (4-5).

If one does not have knowledge of the form of the underlying probability density functions for the x_i, it is not possible to arrive at an optimal estimate of the required parameters. One can use, however, an estimate which is "reasonable."

Define:

$N_1 \triangleq$ Number of patterns in a training set belonging to class 1.

$N_2 \triangleq$ Number of patterns in a training set belonging to class 2.

$n_{1_i} \triangleq$ Number of patterns in a training set belonging to class 1 for which $x_i = 1$.

$n_{2_i} \triangleq$ Number of patterns in a training set belonging to class 2 for which $x_i = 1$.

Reasonable estimates for p_i, q_i, $p(1)$ and $p(2)$ are:

$$p_i = n_{1_i}/N_1$$

$$q_i = n_{2_i}/N_2$$

$$p(1) = N_1/(N_1 + N_2) \qquad (4\text{-}6)$$

Presuming that the machine described above is to be used as a pattern classifier, a means must be devised to partition E^d and to assign binary values to the resulting sub-spaces.

Recall that the time sequence representing a single cough or artifact was divided into 1024 sample sub-segments and measures were calculated for each of these sub-segments. One way of partitioning E^d (in which the patterns are points) is by the use of a non-probabilistic pattern classifier which acts on each sub-segment's measures individually to arrive at a tentative decision as to the sub-segment's class membership. Since we are interested in a two class partition, we may assign a binary value to the sub-segment which corresponds to the tentative decision made. For patterns not included in the training set one would expect errors in classification. If these binary valued decisions are used

as inputs to the binary valued Bayes machine previously described, we should be able to decrease the number of wrong decisions. Our model, then, assumes the form of a layered machine, the first layer being trained by a non-probabilistic algorithm while training for the final layer is probabilistic in form. The first layer will be called the first stage of the pattern classifier and the second layer will be designated as the second stage.

4.2 Non-Probabilistic Pattern Classifier First Stage

The machines examined for use as the first stage were themselves layered machines. Recall that a layered machine implements a piecewise-linear discriminant function.

The first banks of a layered machine may be thought of as a device which performs a mapping from pattern space to an image space. Consider the case where patterns are not linearly separable in pattern space. It can be shown that for a particular training set a first layer may be found so that the mapped patterns in decision space are linearly separable.[10] For the special case of $R = 2$ the first bank consists of a number of TLU's connected in parallel. Training involves determining the number of TLU's required and the weights associated with each TLU. The second bank, consisting of a single TLU, which has as inputs the outputs of the first bank TLU's, is trained so that its output indicates the correct response for all patterns in the training set.

A linear partition of pattern space is formed by positioning hyperplanes in the space so that the space is divided into cells such that no two patterns of opposite categorization reside in the same cell. A non-redundant partition is defined as a partition with the property that if any one of the partitioning hyperplanes is removed, at least two non-empty cells will merge into one cell.

4.2.1 Parallel Hyperplane First Stage Implementation

Given P hyperplanes which form a non-redundant partition of two finite subsets of patterns vectors, a sufficient condition that the sets mapped into image space be linearly separable is that exactly P + 1 cells formed by the partition be occupied by patterns.[10] A partition of pattern space by P parallel hyperplanes fulfiils this sufficiency condition.

The equation of a hyperplane may be written:

$$\underline{C} \cdot \underline{X} = z \tag{4-7}$$

Given the hyperplane described by equation (4-7) in E^d ($\underline{C} \neq \underline{0}$), it may be stated that $\underline{C}$ is a vector normal to the hyperplane. Any vector $k\underline{C}$ is also normal to the hyperplane ($k \neq 0$). The two vectors of unit length $\underline{C}/|\underline{C}|$, $-\underline{C}/|\underline{C}|$ are the unit normals to the hyperplane. $|\underline{z}|/|\underline{C}|$ is the distance of the hyperplane from the origin.[9] Two hyperplanes are parallel if they have the same unit normal.

We may form a non-redundant partition of pattern space with parallel hyperplanes by the implementation of the following algorithm:

1) Select a unit normal vector, $\underline{N}$.

2) Form the dot product of patterns in the training set with the unit normal.

3) Arrange the dot products obtained in step 2) in order of value, largest first.

4) Scan the ordered dot products to find adjacent products that belong to different classes. Let these dot products (which are distances from the origin along the unit normal) be designated as d(i) and d(i + 1).

5) Place a partitioning hyperplane perpendicular to the unit normal at a distance from the origin in the direction the unit unit normal of:

$$z_i = [d(i) + d(i + 1)]/2$$

6) Repeat steps 4) and 5) until all the adjacent dot products of differing class have been found. Let P be the total number of partitioning hyperplanes.

In terms of discriminant function we may write:

$$g_i(\underline{X}) = (-1)^{k+1} [\underline{N} \cdot \underline{X} - z_i] \qquad (4\text{-}8)$$

$$i = 1, \ldots, P$$

$k \triangleq 0$ if the vector associated with d(1) belongs to class 2

$k \triangleq 1$ if the vector associated with d(1) belongs to class 1

It is seen that the first layer of our machine may be implemented by P TLU's. Training consists of finding the z_i in equation (4-8) and is non-probabilistic in form.

Define:

$$y_i = 1 \text{ if } g_i(\underline{X}) > 0 \qquad (4\text{-}9)$$

$$= -1 \text{ if } g_i(\underline{X}) < 0$$

$$i = 1, \ldots, P$$

The vector $\underline{Y}$, which is the output of the first layer, is the input to the single TLU which comprises the second layer of the machine. Since the P hyperplanes have formed a non-redundant partition of pattern space with exactly P + 1 cells occupied, the vectors $\underline{Y}$ in the training set are linearly separable and the discriminant function which performs the separation may be written:

$$g(\underline{Y}) = \underline{W}\cdot\underline{Y} + w_{P+1} \qquad (4\text{-}10)$$

$$g(\underline{Y}) > 0 \Rightarrow \underline{X}\epsilon \text{ class } 1$$

$$g(\underline{Y}) < 0 \Rightarrow \underline{X}\epsilon \text{ class } 2$$

It can be shown that a $\underline{W}'$ which satisfies equation (4-10) is given by:

$$W_i = 1 \qquad i = 1, 2, \ldots, d$$

$$W_{d+1} = 0 \qquad \text{P odd} \qquad (4\text{-}11)$$

$$= (-1)^{k+1} \qquad \text{P even}$$

where k is as defined in (4-8).

It is interesting to note that for P odd the discriminant function for the second layer TLU implements a majority logic decision.

If the parallel hyperplane partition approach to the layered machine is used, it is necessary to determine an efficient means for choosing a unit normal vector. It would be desirable to position the unit normal vector so that the projections of the training vectors on the normal vector are close together for patterns belonging to the same class and far apart for patterns belonging to different classes.

It is possible to obtain such a vector analytically if the underlying probability density functions are known. The "discriminant analysis"[5] technique used in classical statistics does this for vectors from populations with normal distributions and equal variances.

Under certain assumptions regarding the nature of the data (for example, unimodality of the underlying probability density functions) such a vector may be estimated from experimentally gathered points[13]. However, no matter what method is used to arrive at the unit normal vector, the method has inherent shortcomings. One may visualize any number of cases where this technique would fail to separate even disjoint classes. This failure may be traced to the fact that multidimensional data is effectively being reduced to a single dimension in pattern space with an accompanying loss of information.

Despite the obvious drawbacks the algorithm is attractive because of its simplicity and the assurance that a given training set may be separated by linear surfaces. Preliminary preprocessing of the data indicated that the underlying probability density functions were not unimodal. No effort was made, therefore, to obtain a unit normal vector with the desirable properties ennumerated above. The algorithm was programmed, however, using the difference of the means of the patterns in the two classes as the normal vector. The results are ennumerated in Chapter 5 of this report.

4.2.2 Minimum Distance First Stage Implementation

An alternate means of effecting the required partition of E^d (pattern space) takes advantage of the greater "similarity" of patterns which are members of one class as contrasted to patterns which are members of the other class.

For the purposes of this dissertation we shall define similarity in terms of the Euclidian distance between points.

Define:

$$d(\underline{X}, \underline{P}_i) \triangleq [(\underline{X} - \underline{P}_i) \cdot (\underline{X} - \underline{P}_i)]^{\frac{1}{2}}$$

which is the Euclidian distance between $\underline{X}$ and $\underline{P}_i$. If $d(\underline{X}, \underline{P}_i) <$ $d(\underline{X}, \underline{P}_j)$ we shall say that $\underline{X}$ is more similar to $\underline{P}_i$ than to $\underline{P}_j$.

If the $\underline{P}_i$ are prototype points as defined in the section of the dissertation dealing with the minimum distance classifier with respect

to point sets, a similarity decision may be made using the discriminant functions described by equation (2-17), repeated below for convenience:

$$g_i(\underline{X}) = \underline{X} \cdot \underline{P}_i - \tfrac{1}{2}\,\underline{P}_i \cdot \underline{P}_i \qquad (2\text{-}17)$$

Since (2-17) set equal to 0 defines a hyperplane, a linear partition of E^d has been formed. There is no reason to expect that the partition effected is non-redundant, however, if $i > 2$. If the $\underline{P}_i$ are colinear the partitioning hyperplanes are parallel and the situation is as outlined in Section 4.2.1. If the $\underline{P}_i$ are not colinear, more than a single dimension is required to define the decision regions. We may logically expect, therefore, that we have preserved more of the information contained in the original measures than in the colinear case.

Suppose that our situation is as described for the minimum distance classifier and that we may therefore utilize equation (2-19) as a discriminant function to obtain the binary representation of the sub-segment pattern required as the output of the pattern classifier first stage. Alternatively, we may form linear discriminant functions for each pair of $\underline{P}_i$ from differing classes and obtain a vector [illegible] representation of the pattern, that is, effect a mapping from [illegible] space to a binary valued image space. Such a mapping may be accomplished by a bank of TLU's. If m of the $\underline{P}_i$ are prototype points from class 1 and n of the $\underline{P}_i$ are prototype points from class 2, a total of m x n TLU's will be required. A binary logic network with the first bank TLU outputs as

inputs may be designed which will implement a decision identical to that obtained by the use of the discriminant functions described by equation (2-19). In either case the decision surfaces in pattern space are piecewise linear. The binary logic network may possibly (but not necessarily) be implemented by layered TLU's.

The foregoing discussion has presumed the existence of similarity or clustering. For the data with which this paper is concerned, this presumption appears to be valid. Training for the first stage of the pattern classifier consists of the determination of the $\underline{P}_i$ which will be used as prototype patterns for the minimum distance classifier.

We have defined similarity in terms of closeness of points in the sense of Euclidian distance. A reasonable method for determination of the prototype points is to determine the centroid of points which show the greatest similarity. This requires that we find the regions in pattern space in which the patterns cluster.

Two basic approaches to the solution of this problem are evident. One approach is to a priori specify the boundaries of a cluster relative to the centroid. Training then consists of locating the prototype points so that all, or most, of the training patterns are contained in one of the clusters.

A second approach is to place an upper bound on the number of clusters which will be considered. Arbitrary vectors are assigned as a

first approximation to the centroids of the assumed clusters and training is completed by adjusting the centroids by the application of an iterative algorithm.

If the first approach is chosen, the cluster boundaries will define either hyperspheres or hyperellipsoids depending upon whether the assumed distance of the boundaries from the centroid are equal in all directions or are different along the component axes. In either case a preliminary examination of the data will be necessary to determine the distances to be used.

Algorithms implementing both of the above described approaches were programmed. Descriptions are given in Chapter 5.

4.2.3 Summary of First Stage Characteristics

At this point a brief summary concerning the structure of the first stage of the pattern classifier is in order.

The first structure considered was a layered machine, the first bank of which consisted of TLU's whose hyperplane decision surfaces were parallel. It was shown that the orientation of the hyperplane defined by the second layer TLU is fixed while its distance from the origin is dependent only upon whether the number of first bank TLU's is odd or even and, if the number of the first bank TLU's is even, on the class membership of the training pattern which is furtherest from the origin.

Binary representations of the pattern, $\underline{X}$, are available as the output of the first layer TLU's and as the output of the second layer TLU's. It was noted that a given set of patterns could be separated by this structure, but that second layer decision errors would be expected for patterns not included in the training set.

The second structure considered was a minimum distance classifier with respect to point sets. Training for this structure consisted of estimating the prototype patterns. It was noted that two implementations of the minimum distance classifier were possible. The first utilized the discriminant function given as equation (2-19) while the second consisted of a first bank of TLU's feeding a binary switching circuit. The decisions reached by the first stage were identical for the two structures.

For both the minimum distance and parallel hyperplane classifiers, then, two binary representations of a pattern subsegment are available. The first is a binary valued vector which has as elements the output of a first bank of TLU's. The second representation is a single binary value representing the first stage decision.

These binary representations are the input to the probabilistic second stage of the pattern classifier.

4.3 Pattern Classifier Probabilistic Second Stage

4.3.1 Statistically Independent Measures Model

To this point we have been concerned with the ordered set of measures

obtained from a single 1024 sample subsegment of the original digitized data. Define $\underline{X}^{(j)}$ as a first stage binary representation of the jth subsegment. Suppose a segment contains k subsegments. Define:

$$\underline{Z} = (\underline{X}^{(1)}, \underline{X}^{(2)}, \ldots, \underline{X}^{(k)})$$

Utilizing the logarithmic form of the Bayes discriminant function for the symmetric loss function with R = 2 and assuming the $\underline{X}^{(j)}$ statistically indpendent, we may write:

$$g(\underline{Z}) = \sum_{j=1}^{k} \ln \frac{p(\underline{X}^{(j)}|1)}{p(\underline{X}^{(j)}|2)} + \ln \frac{p(1)}{p(2)}$$

If the elements of $\underline{X}^{(j)}$ are assumed to be statistically independent the following relation applies:

$$g(\underline{Z}) = \sum_{j=1}^{k} \sum_{i=1}^{d} \ln \frac{p(x_i^{(j)}|1)}{p(x_i^{(j)}|2)} + \ln \frac{p(1)}{p(2)} \tag{4-12}$$

Define the subsidiary discriminant functions:

$$g^{(j)}(\underline{X}) = \sum_{i=1}^{d} \ln \frac{p(x_i^{(j)}|1)}{p(x_i^{(j)}|2)} \tag{4-13}$$

Equation (4-13) is of the form of equation (4-1) with p(1) = p(2) (maximum likelihood discriminant).

Substituting (4-13) into (4-12) yields:

$$g(\underline{Z}) = \sum_{j=1}^{k} g^{(j)}(\underline{X}) + \ln \frac{p(1)}{p(2)} \tag{4-14}$$

Extending the nomenclature previously used to define the various probabilities, we define:

$$p(x_i^{(j)} = 1 \mid 1) \triangleq p_i^{(j)}$$

$$p(x_i^{(j)} = 0 \mid 1) \triangleq 1 - p_i^{(j)}$$

$$p(x_i^{(j)} = 1 \mid 2) \triangleq q_i^{(j)}$$

$$p(x_i^{(j)} = 0 \mid 2) \triangleq 1 - q_i^{(j)}$$

Then:

$$g^{(j)}(\underline{X}) = \sum_{i=1}^{d} x_i \,\ell n \,\frac{p_i^{(j)}\,(1 - q_i^{(j)})}{q_i^{(j)}\,(1 - p_i^{(j)})} + \sum_{i=1}^{d} \ell n \,\frac{(1 - p_i^{(j)})}{(1 - q_i^{(j)})} \tag{4-15}$$

Substitution of (4-15) into (4-14) yields:

$$g(\underline{Z}) = \sum_{j=1}^{k} \sum_{i=1}^{d} x_j \,\ell n \,\frac{p_i^{(j)}(1 - q_i^{(j)})}{q_i^{(j)}(1 - p_i^{(j)})} + \sum_{j=1}^{k} \sum_{i=1}^{d} \ell n \,\frac{(1 - p_i^{(j)})}{(1 - q_i^{(j)})} + \ell n \,\frac{p(1)}{p(2)} \tag{4-16}$$

d = dimension of $\underline{X}$

k = No. subsegments

Equation (4-16) is of the form:

$$g(\underline{Z}) = \underline{W} \cdot \underline{Z} + w_{\ell+1} \tag{4-17}$$

$\underline{W}$ and $\underline{Z}$ ℓ dimensional

$\ell = d \times k$

where

$$w_m = w_{j\times d+i} = \ln \frac{p_i^{(j)}(1 - q_i^{(j)})}{q_i^{(j)}(1 - p_i^{(j)})}$$

$j = 1, \ldots, k$

$i = 1, \ldots, d$ (4-18)

$m = 1, 2, \ldots, \ell$

$\ell = d \times k$

$$w_{\ell+1} = \sum_{j=1}^{k} \sum_{i=1}^{d} \frac{(1 - p_i^{(j)})}{(1 - q_i^{(j)})} + \ln \frac{p(1)}{1 - p(1)}$$

Training then consists of estimating the $p_i^{(j)}$, $q_i^{(j)}$ and $p(1)$, $(i = 1, \ldots, d ; j = 1, \ldots, k)$.

4.3.2 Markoff l Distributed Measures

The presumption of statistical indpendence invoked in Section 4.3.1 was justified on the basis of computational managability. If the elements of the binary valued vector represent time sequential first stage decisions, a possibly better assumption is that the measures comprise a Markoff chain.

The general logarithmic form of the two class symmetric loss Bayes discriminant function is repeated below:

$$g(\underline{X}) = \ln p(\underline{X} \mid 1) - \ln p(\underline{X} \mid 2) + \ln \frac{p(1)}{p(2)} \tag{4-19}$$

$$p(\underline{X} \mid j) = p(x_1, x_2, \ldots, x_k \mid j) \tag{4-20}$$

$$j = 1, 2$$

k = number of sequential first stage decisions

If $(x_1, x_2, \ldots, x_k)$ is a Markoff chain, by definition[11]:

$$p(x_i \mid x_{i-1}, x_{i-2}, \ldots, x_1) = p(x_i \mid x_{i-1}) \tag{4-21}$$

$$i = 2, 3, \ldots, k$$

Then, by the definition of conditional probability:

$$\begin{aligned} p(x_k, x_{k-1}, \ldots, x_1) &= p(x_k \mid x_{k-1}) p(x_{k-1}, x_{k-2}, \ldots, x_1) \\ &= p(x_k \mid x_{k-1})\, p(x_{k-1} \mid x_{k-2})\, p(x_{k-2}, \ldots, x_1) \\ &\ldots \end{aligned}$$

$$p(x_k, x_{k-1}, \ldots, x_1) = p(x_1) \prod_{i=2}^{k} p(x_i \mid x_{i-1}) \tag{4-22}$$

Substitution of (4-22) into equation (4-19) yields:

$$\begin{aligned} g(\underline{X}) = {} & \ln \frac{p(x_1 \mid 1)}{p(x_1 \mid 2)} + \sum_{i=2}^{k} \ln \frac{p(x_i \mid 1, x_{i-1})}{p(x_i \mid 2, x_{i-1})} \\ & + \ln \frac{p(1)}{1 - p(1)} \end{aligned} \tag{4-23}$$

Recall that $\underline{X}$ may take on values of only 1 and 0. Define:

$$p(x_1 = 1 \mid 1) \triangleq p$$

$$p(x_i = 1 \mid 1, x_{i-1} = 1) \triangleq r_i$$

$$p(x_i = 1 \mid 1, x_{i-1} = 0) \triangleq s_i$$

$$
\begin{aligned}
&p(x_i = 0 \mid 1, x_{i-1} = 1) \triangleq 1 - r_i \\
&p(x_i = 0 \mid 1, x_{i-1} = 0) \triangleq 1 - s_i \\
&p(x_1 = 1 \mid 2) \triangleq q \\
&p(x_i = 1 \mid 2, x_{i-1} = 1) \triangleq t_i \\
&p(x_i = 1 \mid 2, x_{i-1} = 0) \triangleq v_i \\
&p(x_i = 0 \mid 2, x_{i-1} = 1) \triangleq 1 - t_i \\
&p(x_i = 0 \mid 2, x_{i-1} = 0) \triangleq 1 - v_i \\
&\qquad\qquad\qquad\qquad i = 2, \ldots, k
\end{aligned}
\tag{4-24}
$$

Making use of the fact that $x_i = 0$ or $x_i = 1$ we may write:

$$
\begin{aligned}
\ln \frac{p(x_i \mid 1)}{p(x_i \mid 2)} = \\
& x_i \left[x_{i-1} \ln \frac{r_i}{t_i} + (1 - x_{i-1}) \ln \frac{s_i}{v_i} \right] \\
& + [1 - x_i] \left[x_{i-1} \ln \frac{1-r_i}{1-t_i} + (1 - x_{i-1}) \ln \frac{1-s_i}{1-v_i} \right] \\
& = x_i \ln \frac{s_i(1-v_i)}{v_i(1-s_i)} \\
& + x_i x_{i-1} \ln \frac{r_i v_i (1-t_i)(1-s_i)}{t_i s_i (1-r_i)(1-v_i)} \\
& + x_{i-1} \ln \frac{(1-r_i)(1-v_i)}{(1-t_i)(1-s_i)} \\
& + \ln \frac{1-s_i}{1-v_i}
\end{aligned}
\tag{4-25}
$$

Also:

$$\ln \frac{p(x_1|1)}{p(x_1|2)} = x_1 \ln \frac{p}{q} + (1-x_1) \ln \frac{(1-p)}{(1-q)}$$

$$= x_1 \ln \frac{p(1-q)}{q(1-p)} + \ln \frac{(1-p)}{(1-q)} \qquad (4\text{-}26)$$

Substitution of equations (4-25) and (4-26) into equation (4-23) yields:

$$g(\underline{X}) = x_1 \ln \frac{p(1-q)}{q(1-p)} + \sum_{i=2}^{k} x_i \ln \frac{s_i(1-v_i)}{v_i(1-s_i)}$$

$$+ \sum_{i=2}^{k} x_i\, x_{i-1} \ln \frac{r_i v_i (1-t_i)(1-s_i)}{t_i s_i (1-r_i)(1-v_i)}$$

$$+ \sum_{i=2}^{k} x_{i-1} \ln \frac{(1-r_i)(1-v_i)}{(1-t_i)(1-s_i)} + \sum_{i=2}^{k} \ln \frac{1-s_i}{1-v_i} \qquad (4\text{-}27)$$

$$+ \ln \frac{1-p}{1-q} + \ln \frac{p(1)}{1-p(1)}$$

If k is fixed, equation (4-27) may be written as:

$$g(\underline{X}) = x_1 \ln \frac{p(1-q)(1-r_2)(1-v_2)}{q(1-p)(1-t_2)(1-s_2)}$$

$$+ \sum_{i=2}^{k-1} x_i \ln \frac{s_i(1-v_i)(1-r_{i+1})(1-v_{i+1})}{v_i(1-s_i)(1-t_{i+1})(1-s_{i+1})}$$

$$+ x_k \ln \frac{s_k(1-v_k)}{v_k(1-s_k)}$$

(Equation cont'd on next page)

$$+ \sum_{i=2}^{k} x_i \, x_{i-1} \, \ln \frac{r_i \, v_i \, (1-t_i)(1-s_i)}{t_i \, s_i \, (1-r_i)(1-v_i)}$$

$$+ \sum_{i=2}^{k} \ln \frac{(1-s_i)}{(1-v_i)} + \ln \frac{1-p}{1-q} + \ln \frac{p(1)}{1-p(1)} \quad (4-28)$$

Equation (4-28) is of the form:

$$g(\underline{X}) = \underline{W} \cdot \underline{X} + \underline{A} \cdot \Phi(\underline{X}) + w_{k+1} \qquad (4-29)$$

where:

$$\varphi_{j-1}(\underline{X}) = x_j \, x_{j-1} \qquad j = 2, \ldots, k \qquad (4-30)$$

$$\varphi_k(\underline{X}) = 0$$

Since x is binary valued:

$$\begin{aligned} x_j \, x_{j-1} &= 0 \qquad \text{for } x_j \neq x_{j-1} \neq 1 \\ x_j \, x_{j-1} &= 1 \qquad \text{for } x_j = x_{j-1} = 1 \end{aligned} \qquad (4-31)$$

Equation (4-31) may be implemented by a TLU; equation (4-29) may therefore be implemented by a layered machine. Since a layered machine implements a piecewise linear discriminant function, $g(\underline{X})$ in equation (4-29) is piecewise linear. If this realization is used as the second stage of a pattern classifier for which the first stage is piecewise linear (for example the first stage described in Section 4.2 of this paper), the composite machine is piecewise linear.

Training for the Markoff machine consists of estimating the quantities defined in (4-24).

Define:

P = Number of class 1 training patterns for which $x_1 = 1$.

R_i = Number of class 1 training patterns for which $x_i = 1$ and $x_{i-1} = 1$; $i \quad 2, 3, \ldots, m$.

S_i = Number of class 1 training patterns for which $x_i = 1$ and $x_{i-1} = 0$; $i = 2, 3, \ldots, m$.

Q = Number of class 2 training patterns for which $x_1 = 1$.

T_i = Number of class 2 training patterns for which $x_i = 1$ and $x_{i-1} = 1$; $i = 2, 3, \ldots, m$.

V_i = Number of class 2 training patterns for which $x_i = 1$ and $x_{i-1} = 0$; $i = 2, 3, \ldots, m$.

N_{1j} = Number of class 1 training patterns having at least j components (assuming that the patterns can have a varying number of components); $j = 1, 2, \ldots, m$. (4-32)

N_{2j} = Number of class 2 training patterns having at least j components; $j = 1, 2, \ldots, m$.

M_{1i} = Number of class 1 training patterns for which $x_{i-1} = 1$; $i = 2, 3, \ldots, m$.

M_{2i} = Number of class 2 training patterns for which

$x_{i-1} = 1$, $i = 2, 3, \ldots, m$.

m = Minimum of the maximum number of components in class 1 training patterns or class 2 training patterns.

Reasonable estimates for the quantities defined in (4-24) are:

$$
\begin{aligned}
p &= P/N_{11} & q &= Q/N_{21} \\
r_i &= R_i/M_{1i} & t_i &= T_i/M_{2i} \\
s_i &= S_i/(N_{1i}-M_{1i}) & v_i &= V_i/(N_{2i}-M_{2i})
\end{aligned}
\qquad (4\text{-}33)
$$

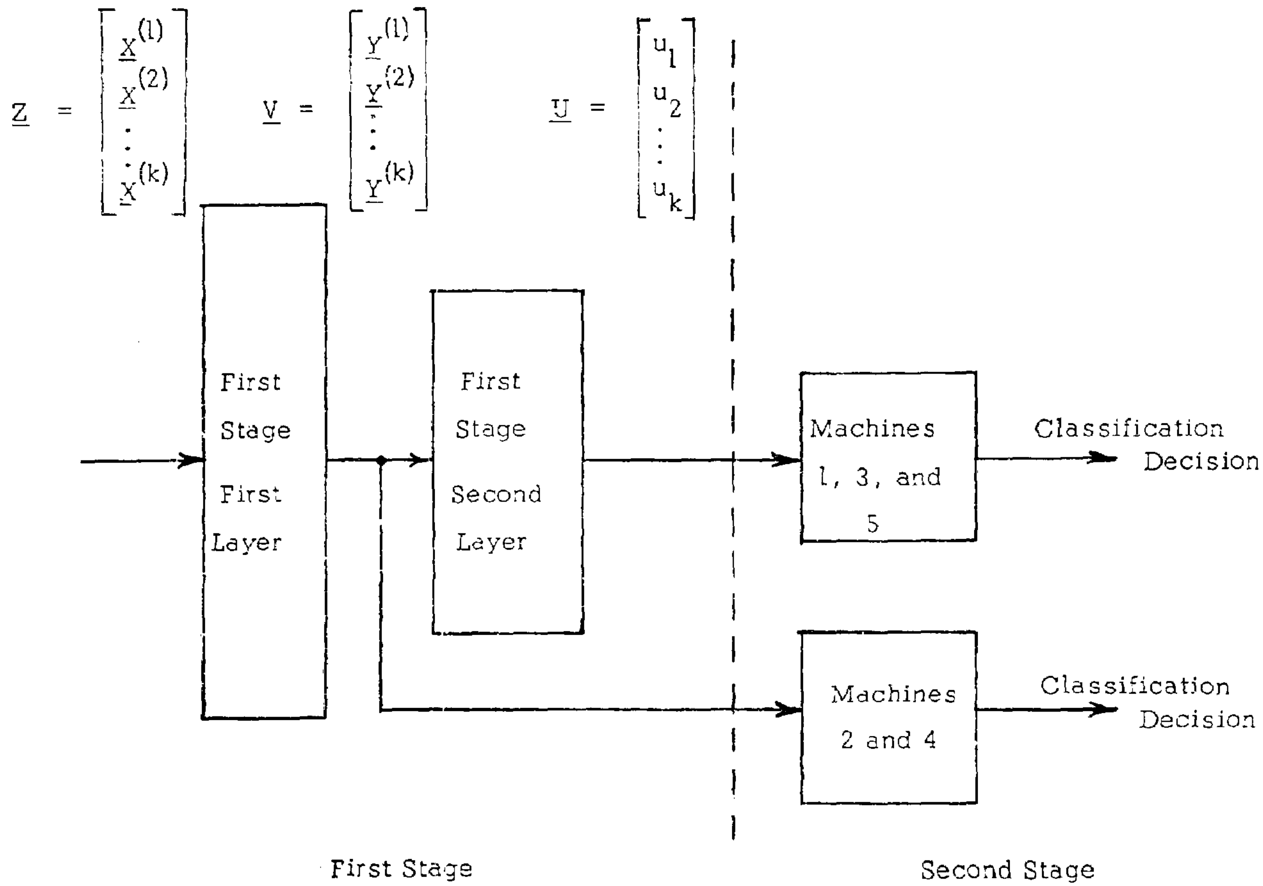
$$i = 2, 3, \ldots, m$$

If k, the number of components in a pattern, is fixed, equation (4-28) may be used as the recognition discriminant function. If k is variable from pattern to pattern it is necessary to use equation (4-27) to effect recognition since it is not possible to compute the constant term until the dimensionality of the pattern is known.

4.4 Composite Machine Structure

Recall that the first stage of the pattern classifier described in Section 4.2 maps the pattern into a binary representation. These binary values may be either 0 or 1 or +1 and -1, depending upon the representation required for the succeeding stage.

Figure 4-1 is a block diagram of the pattern classifiers implemented.



Pattern Classifier
Block Diagram

Figure 4-1

Referring to this figure, $\underline{X}^{(i)}$ is a vector consisting of d real valued measures for the ith subsegment in a pattern. There are a total of k subsegments in a pattern. $\underline{Z}$ is therefore a dxk dimensional vector which represents a complete pattern.

Assume that the first stage of the first layer of the pattern classifier consists of p TLU's. $\underline{Y}^{(i)}$ is a binary valued p dimensional vector representing the ith subsegment of the pattern. $\underline{V}$ is the first stage--first layer binary representation of a total pattern and consists of pxk ordered binary values.

The output of the second layer of the first stage of the pattern classifier is a binary valued scalar u_i which represents the preliminary decision made by the first stage of the pattern classifier as to the classification of the ith subsegment of a pattern. $\underline{U}$ is a k dimensional binary valued vector which represents the preliminary sequential classification decisions of the first stage of the pattern classifier for a complete pattern.

Regardless of the form that the pattern classifier takes, errors in classification are to be expected. It is logical to assume that an increase in the complexity of the machine chosen would be reflected in a lowering of the probability of erroneous classification.

Five second stage classifiers were implemented. The differences in machine complexity will be apparent from the descriptions of the

machines. The structures of the machines follow from the presumption made concerning the statistical properties of the first stage binary vectors. These presumptions are outlined in the sections of this chapter of the dissertation which follow.

4.4.1 Machine Number 1

Referring to Figure 4-1, machine number 1 operated on $\underline{U}$, the first stage--second layer output. This implementation was made under the assumption that the u_i's were statistically independent and identically distributed for $i = 1, 2, \ldots, k$. The applicable discriminant function is given below as equation (4-34):

$$g_1(\underline{U}) = \sum_{i=1}^{k} u_i \, \ell n \, \frac{p(1-q)}{q(1-p)} + k \, \ell n \, \frac{1-p}{1-q} + \frac{p(1)}{1-p(1)} \tag{4-34}$$

p and q were estimated from a training subset without regard to subsegment number. It is seen that only three values must be stored for the second stage classifier.

4.4.2 Machine Number 2

Referring to Figure 4-1, machine number 2 operated on $\underline{V}$, the first stage--first bank binary representation. This implementation was made under the presumption that the $y_j^{(i)}$'s were statistically independent over $j = 1, 2, \ldots, p$ and identically distributed for $i = 1, 2, \ldots, k$. The applicable discriminant function is given below as equation (4-35):

$$g_2(\underline{V}) = \sum_{i=1}^{k} \sum_{j=1}^{p} y_j^{(i)} \ln \frac{p_j(1-q_j)}{q_j(1-p_j)}$$

$$+ k \sum_{j=1}^{p} \ln \frac{(1-p_j)}{(1-q_j)} + \ln \frac{p(1)}{1-p(1)} \qquad (4-35)$$

The p_j and q_j were estimated from a training subset without regard to subsegment number.

4.4.3 Machine Number 3

Referring to figure 4-1, machine number 3 operated on $\underline{U}$, the first stage-second layer output. This implementation was made under the assumption that the u_i's were statistically independent, but not necessarily identically distributed for $i = 1, 2, \ldots, k$. The applicable discriminant function is given below as equation (4-36):

$$g_3(\underline{U}) = \sum_{i=1}^{k} u_i \ln \frac{p_i(1-q_i)}{q_i(1-p_i)} + \sum_{i=1}^{k} \ln \frac{(1-p_i)}{(1-q_i)}$$

$$+ \ln \frac{p(1)}{1-p(1)} \qquad (4-36)$$

The p_i and q_i were estimated from a training subset using only the ith subsegments to estimate p_i and q_i.

4.4.4 Machine Number 4

Referring to Figure 4-1, machine number 4 operated on $\underline{V}$, the first stage--first bank binary representation. This implementation was made under the presumption that the $y_j^{(i)}$'s were statistically independent

for $i = 1, 2, \ldots, k$; $j = 1, 2, \ldots, p$, but not necessarily identically distributed over i. The applicable discriminant function is given below as equation (4-37):

$$g_4(\underline{V}) = \sum_{i=1}^{k} \sum_{j=1}^{p} \left[y_j^{(i)} \ln \frac{p_j^{(i)} (1-q_j^{(i)})}{q_j^{(i)} (1-p_j^{(i)})} + \ln \frac{(1-p_j^{(i)})}{(1-q_j^{(i)})} \right] + \ln \frac{p(1)}{1-p(1)} \tag{4-37}$$

$p_j^{(i)}$ and $q_j^{(i)}$ were estimated using only jth first bank TLU output for the ith subsegment in the training subset.

4.4.5 Machine Number 5

Machine number 5 is identical to machine number 3 except it was assumed that the u_i's were Markoff 1 distributed rather than being statistically independent. The applicable discriminant function is given below as equation (4-38):

$$\begin{aligned} g_5(\underline{U}) = {} & u_1 \ln \frac{p(1-q)(1-r_2)(1-v_2)}{q(1-p)(1-t_2)(1-s_2)} \\ & + \sum_{i=2}^{k-1} u_i \ln \frac{s_i(1-v_i)(1-r_{i+1})(1-v_{i+1})}{v_i(1-s_i)(1-t_{i+1})(1-s_{i+1})} \\ & + u_k \ln \frac{s_k(1-v_k)}{v_k(1-s_k)} \end{aligned} \tag{4-38}$$

(equation continued on next page)

$$+\sum_{i=2}^{k} u_i\, u_{i-1}\, \ell n\, \frac{r_i v_i (1-t_i)(1-s_i)}{t_i s_i (1-r_i)(1-v_i)}$$

$$+\sum_{i=2}^{k} \ell n\, \frac{(1-s_i)}{(1-v_i)} + \ell n\, \frac{1-p}{1-q} + \ell n\, \frac{p(1)}{1-p(1)}$$

The various probabilities were estimated by use of the relationships given in (4-32) and (4-33) with respect to u.

CHAPTER V

RESULTS

5.1 Data Processed

As outlined in section 3.2 of this report, data processed was from one of two categories. These categories were:

Category one -- data recorded at Woodlawn Hospital

Category two -- simulated coughs and artifacts recorded at The University of Texas at Austin.

Category one data was taken from recordings made without benefit of the start-stop recording procedure outlined in section 3.2.1. Rather, the original audio tape recording was edited and re-recorded prior to digitization. Included in this type of data were forty-one cough segments (patterns) which contained a total of 444 sub-segments and twenty-eight artifact segments which contained a total of 342 sub-segments (a total of sixty-nine segments and 786 sub-segments).

Category two data included 112 cough segments, which contained 450 sub-segments and twenty-eight artifact segments which contained 850 sub-segments, a total of 140 segments and 1350 sub-segments.

When category one data was used for training, it was possible to estimate the a priori probability of occurrence of the cough and artifact classes. In the case of category two data, such an estimation was not

possible, since the true relative number of occurrences of cough and artifact patterns was not preserved. The a priori probabilities for this latter case were set equal, which resulted in maximum likelihood decisions rather than Bayesian decisions.

The discrepancy between the lengths of the category one and category two data signals was due to the recording techniques employed. The background noise in category one data tended to make the coughs longer since the individual "hacks" which constituted a cough series were placed in the same segment in many cases. The category two artifact segments tend to contain many sub-segments because they were recorded from continuous commerical recording signals.

The differing lengths of signal did not have a significant effect on training and classification results, however, since it was required that sufficient numbers of sub-segments be available from both the cough and artifact classes to obtain reasonable estimates of the probabilities involved. During the classification phase, sub-segments in excess of this number were ignored.

5.2 Measure Calculation

Measures were calculated by the procedures outlined in section 3.3 of this paper.

Computation of the measures for a 1024 point sub-segment required approximately 0.8 second, of which approximately 0.7 second was expended on calculation of the Fourier series coefficients. The measure numbers to

which references are made in the discussion which follows are listed in Table 5-3.

Prior to utilization of measure number 139 (average zero-crossing frequency) and measures 140 through 228 (spectral analysis measures), it was necessary to normalize the calculated values so that the resultant measures had a nominal maximum value of one. Failure to effect this normalization would tend to place emphasis on those measures with the largest magnitude.

The empirical normalizing relationships used were as follows (v is the calculated value of the measure):

Measure No. 139 (average zero-crossing frequency) --$(\log_{10} v)/16500$

Measure Nos. 140-148 (magnitude squared Fourier coefficient bands) - $8.33 \times 10^{-3} \log_{10} v$

Measure Nos. 149-168 (frequency at which spectral peaks occurred) and measure Nos. 189-228 (spectral peak 1/4 power frequencies) - $0.025 \log_{10} v$

Measure Nos. 169-188 (magnitude squared of spectral peaks) - $0.01 \log_{10} v$.

5.3 Measure Selection Results

The procedure outlined in section 3.4 was applied to category one data and category two data separately. Table 5-1 is a list of the results obtained for the fifty highest ranked measures for category one data and

Table 5-2 for category two data. The criterion value in columns three and six of the tables is an estimate of $\int_{-\infty}^{\infty} |p(v/1) - p(v/2)| dv$ where p(v/1) and p(v/2) are the marignal probability functions conditioned on the measure v having originated from a signal belonging to the cough and artifact classes respectively. The measure numbers in columns two and four are identified in Table 3-5. It will be noted that a maximum possible criterion value of 2.0 would indicate that the classes were disjoint with respect to that particular measure.

The measures listed in Table 5-1 will be referred to as "measure set A." The measures listed in Table 5-2 will be referred to as "measure set C." The first twenty-five of the measures listed in Table 5-2 will be referred to as "measure set D."

Because category one data contained considerable noise, it would be expected that the spectral features would be masked to some extent. This conclusion is verified by the feature selection results. For this group of data, the features emphasized are the envelope difference and zero-crossing features, while the spectral features are almost entirely excluded.

In the case of category two data, for which the signal was relatively clean, spectral features predominate as those most likely to aid in recognition. The amplitude density and envelope amplitude density features are also stressed.

The estimated absolute difference in marginal densities indicates that recognition error rates should be considerably lower than those for

Table 5-1

Measure Selection Results

Category One Data

Rank	Feature Number	Criterion Value	Rank	Feature Number	Criterion Value
1	209	0.727	18	191	0.377
2	149	0.721	19	12	0.377
3	189	0.701	20	151	0.374
4	138	0.672	21	94	0.364
5	48	0.640	22	137	0.363
6	135	0.626	23	102	0.362
7	136	0.622	24	106	0.360
8	97	0.466	25	40	0.359
9	92	0.466	26	87	0.356
10	144	0.451	27	93	0.348
11	47	0.447	28	105	0.345
12	49	0.414	29	18	0.333
13	46	0.392	30	218	0.333
14	96	0.387	31	104	0.332
15	211	0.386	32	17	0.329
16	103	0.384	33	158	0.327
17	90	0.384	34	107	0.324

Rank	Feature Number	Criterion Value
35	198	0.321
36	10	0.317
37	139	0.314
38	100	0.314
39	85	0.310
40	50	0.310
41	44	0.305
42	91	0.301
43	11	0.300
44	99	0.299
45	51	0.296
46	88	0.291
47	24	0.289
48	95	0.283
49	108	0.283
50	45	0.282

Table 5-2

Measure Selection Results

Category Two Data

Rank	Feature Number	Criterion Value	Rank	Feature Number	Criterion Value
1	11	1.124	18	176	0.920
2	12	1.105	19	173	0.913
3	17	1.099	20	20	0.913
4	16	1.070	21	48	0.911
5	21	1.068	22	175	0.911
6	22	1.048	23	174	0.908
7	10	1.007	24	179	0.903
8	49	0.984	25	23	0.897
9	148	0.942	26	170	0.896
10	177	0.940	27	186	0.892
11	13	0.938	28	146	0.889
12	178	0.938	29	185	0.889
13	171	0.936	30	184	0.887
14	147	0.933	31	180	0.883
15	142	0.927	32	181	0.879
16	169	0.927	33	187	0.877
17	172	0.925	34	183	0.877

Rank	Feature Number	Criterion Value
35	188	0.868
36	182	0.860
37	9	0.849
38	144	0.849
39	189	0.836
40	24	0.825
41	145	0.802
42	47	0.800
43	8	0.791
44	54	0.791
45	25	0.726
46	50	0.721
47	41	0.709
48	43	0.702
49	42	0.684
50	7	0.645

category one data. Results presented in section 5.5 show that such is indeed the case.

An additional set of measures was selected on the basis of simplicity of implementation of the pattern classifier by special purpose circuitry. Zero-crossing measures and envelope difference measures were promising candidates in the light of the measure selection results for category one data. The twenty-eight measures chosen were measures number 85 through 198 and measure numbers 135 through 138. These measures will be referred to as "measure set B."

5.4 Training Set Selection

The first stage of the pattern classifier was trained without regard to the location of a sub-segment within a segment. The second stage of the classifier, on the other hand, took the order in which the sub-segments occur into account. Two types of training classes were therefore required; the first consisted of unordered sub-segments and the second of complete segments (patterns).

Training and classification were made for each of the two categories of data separately. A set of training patterns was therefore chosen from each of the categories.

The procedure used for selecting the training sets is given below:

1) A specified percentage of the patterns (segments) belonging to each of the classes (artifact and cough) were selected at

random. These constituted the training set and were used to train the second stage of the pattern classifier.

2) A specified percentage of the sub-segments contained in each of the patterns selected in 1) were chosen at random to form a training sub-set. This sub-set of training sub-segments was used for training the first stage of the pattern classifier.

The relative frequency of occurrence of coughs and artifacts is preserved by the above training set selection. Use of the training set to estimate a priori probability of occurrence is therefore valid (for category one data).

Seventy-five per cent of the available category one patterns were selected for the training set. This set consisted of thirty-one cough patterns and twenty-one artifact patterns, a total of fifty-two patterns (out of an available sixty-nine).

Fifty per cent of the sub-segments from each of the training patterns were selected for the training sub-set. The sub-set consisted of 164 cough sub-segments and 132 artifact sub-segments, a total of 296 sub-segments.

The category two data training set consisted of fifty per cent of the available patterns. The set contained fifty-five cough patterns and fourteen artifact patterns, a total of sixty-nine patterns.

The training sub-set for category two data was made up of 50% of the sub-segments from each of the training patterns - 108 cough

sub-segments and 206 artifact sub-segments, a total of 314 sub-segments.

5.5 Machine Training and Classification Results

Training and recognition using category one data were accomplished for four first stage configurations used in conjunction with four second stage configurations; the Markoff-1 machine was not implemented. Training and recognition using category two data were conducted for the first stage configuration chosen as a final first stage and for all five second stage configurations. The second stage implementations were as described in sections 4.4.1 through 4.4.5.

The first stage configurations utilized are listed below:

1) Parallel hyperplane implementation (as described in section 4.2.1).
2) Modified parallel hyperplane implementation
3) Minimum distance with respect to point sets configuration with pre-set cluster boundaries (described in section 4.2.2)
4) Minimum distance with respect to point sets configuration with boundaries determined by an iterative algorithm (described in section 4.2.2).

The first stage configuration listed as 4) above was chosen as a final first stage for the pattern classifier.

Training and recognition results are tabulated in Tables 5-3 through 5-14. The following abbreviations were used:

A ---- Artifact Class

C ---- Cough Class

B/D -- a decision was made that the pattern belonged to class B when in actuality it was a member of class D.

B = A, C ; D = A, C

The data identification and measure set nomenclature previously defined are repeated below for convenience:

Category one data --- recorded at Woodlawn Hospital

Category two data --- simulated coughs and artifacts

Measure set A --- the fifty measures described in Table 5-1

Measure set B --- the twenty-eight measures described in the last paragraph of section 5.3

Measure set C --- the fifty measures described in Table 5-2

Measure set D --- the first twenty-five measures described in Table 5-2.

5.5.1 Preliminary Configuration Results

Tables 5-3 and 5-4 are tabulations of the training and pattern classification results for the parallel hyperplane first stage configuration. For reasons which were previously outlined, it was not expected that this configuration should perform well for patterns not included in the training set. The tabulation of the first stage-second layer decisions for sub-segments (excluding the training sub-set) indicates that this was the case.

Table 5-3

Pattern Classifier Results
Parallel Hyperplane First Stage Implementation

Category One Data -- Measure Set A

Measure Set A (50 measures category one data)
Number of First Stage Hyperplanes --109

First stage - second layer decisions:

	C/C	A/C	C/A	A/A
Excluding Training sub-set	166	114	101	109
Training sub-set	164	0	0	132
Total	330	114	101	241

Second stage decisions:

Machine No.	C/C	A/C	C/A	A/A	Patterns Considered
1	10	0	6	1	Exc. training set
2	7	3	5	2	"
3	9	1	3	4	"
4	7	3	3	4	"
1	30	1	5	16	Training set
2	28	3	12	9	"
3	30	1	0	21	"
4	28	3	5	16	"
1	40	1	11	17	All
2	35	6	17	11	"
3	39	2	3	25	"
4	35	6	8	20	"

Table 5-4

Pattern Classifier Results
Parallel Hyperplane First Stage Implementation

Category One Data -- Measure Set E

Number of First Stage Hyperplanes -- 111

First stage - second layer decisions:

	C/C	A/C	C/A	A/A
Excluding Training sub-set	200	80	109	101
Training sub-set	164	0	0	132
Total	364	80	109	233

Second stage decisions:

Machine No.	C/C	A/C	C/A	A/A	Patterns Considered
1	10	0	6	1	Excl. Training set
2	7	3	4	3	"
3	10	0	4	3	"
4	7	3	3	4	"
1	31	0	5	16	Training set
2	26	5	6	15	"
3	31	0	1	20	"
4	29	2	5	16	"
1	41	0	11	17	All
2	33	8	10	18	"
3	41	0	5	23	"
4	36	5	8	20	"

Tabi. 5-5

Pattern Classifier Results
Modified Parallel Hyperplane First Stage Implementation

Category One Data -- Measure Set A

Number of First Stage Hyperplanes -- 9

First stage - second layer decisions:

	C/C	A/C	C/A	A/A
Excluding Training sub-set	204	76	118	92
Training sub-set	144	20	46	86
Total	348	96	164	178

Second stage decisions:

Machine No.	C/C	A/C	C/A	A/A	Patterns Considered
1	8	2	5	2	Excl. training set
2	8	2	5	2	"
3	7	3	5	2	"
4	6	4	4	3	"
1	29	2	11	10	Training set
2	26	5	12	9	"
3	28	3	9	12	"
4	28	3	6	15	"
1	37	4	16	12	All
2	34	7	17	11	"
3	35	6	14	14	"
4	34	7	10	18	"

Table 5-6

Pattern Classifier Results
Modified Parallel Hyperplane First Stage Implementation

Category One Data -- Measure Set B

Number of First Stage Hyperplanes -- 9

First stage - second layer decisions:

	C/C	A/C	C/A	A/A
Excluding Training sub-set	217	63	130	80
Training sub-set	140	24	38	94
Total	357	87	168	174

Second stage decisions:

Machine No.	C/C	A/C	C/A	A/A	Patterns Considered
1	9	1	5	2	Excl. training set
2	9	1	4	3	"
3	9	1	5	2	"
4	8	2	3	4	"
1	27	4	6	15	Training set
2	27	4	7	14	"
3	27	4	4	17	"
4	29	2	5	16	"
1	36	5	11	17	All
2	36	5	11	17	"
3	36	5	9	19	"
4	37	4	8	2	"

Table 5-7

Pattern Classifier Results
Fixed Radius Hypersphere Clustering First Stage Implementation

Category One Data -- Measure Set A

Number of first stage prototype points:
Cough - 5 Artifact - 6

First stage - second layer decisions:

	C/C	A/C	C/A	A/A
Excluding Training sub-set	178	102	63	147
Training sub-set	109	55	45	87
Total	287	157	108	234

Second stage decisions:

Machine No.	C/C	A/C	C/A	A/A	Patterns Considered
1	9	1	4	3	Excl. training set
2	8	2	5	2	"
3	9	1	4	3	"
4	9	1	2	5	"
1	26	5	5	16	Training set
2	28	3	14	7	"
3	26	5	4	17	"
4	30	1	4	17	"
1	35	6	9	19	All
2	36	5	19	9	"
3	35	6	8	20	"
4	39	2	6	22	"

Table 5-8

Pattern Classifier Results
Fixed Radius Hypersphere Clustering First Stage Implementation

Category One Data -- Measure Set B

Number of first stage prototype points:
Cough - 4 Artifact - 8

First stage - second layer decisions:

	C/C	A/C	C/A	A/A
Excluding Training sub-set	190	90	48	152
Training sub-set	75	89	40	92
Total	265	179	88	254

Second stage decisions:

Machine No.	C/C	A/C	C/A	A/A	Patterns Considered
1	10	0	4	3	Excl. Training set
2	10	0	2	5	"
3	10	0	4	3	"
4	9	1	5	2	"
1	25	6	8	13	Training set
2	28	3	14	7	"
3	25	6	8	13	"
4	28	3	3	18	"
1	35	6	12	16	All
2	38	3	16	12	"
3	35	6	12	16	"
4	37	4	8	20	"

Table 5-9

Pattern Classifier Results
Final Machine Configuration*

Category One Data -- Measure Set A

Number of first stage prototype points:
Cough - 18 Artifact - 13

First stage - second layer decisions:

	C/C	A/C	C/A	A/A
Excluding Training sub-set	222	58	77	133
Training sub-set	151	13	35	97
Total	373	71	112	230

Second stage decisions:

Machine No.	C/C	A/C	C/A	A/A	Patterns Considered
1	8	2	4	3	Excl. training set
2	9	1	5	2	"
3	7	3	4	3	"
4	8	2	2	5	"
5	6	4	3	4	"
1	30	1	5	16	Training set
2	29	2	15	6	"
3	28	3	3	18	"
4	30	1	4	17	"
5	31	0	3	18	"
1	38	3	9	19	All
2	38	3	20	8	"
3	35	6	7	21	"
4	38	3	6	22	"
5	37	4	6	22	"

* First stage prototype points computed by an iterative algorithm.

Table 5-10

Pattern Classifier Results
Final Machine Configuration

Category One Data -- Measure Set B

Number of first stage prototype points:
Cough - 16 Artifact - 15

First stage - second layer decisions:

	C/C	A/C	C/A	A/A
Excluding Training sub-set	210	70	44	166
Training sub-set	137	27	29	103
Total	347	97	73	269

Second stage decisions:

Machine No.	C/C	A/C	C/A	A/A	Patterns Considered
1	9	1	3	4	Excl. training set
2	9	1	5	2	"
3	9	1	0	7	"
4	9	1	4	3	"
5	9	1	0	7	"
1	26	5	10	11	Training set
2	27	4	7	14	"
3	26	5	5	16	"
4	28	3	4	17	"
5	29	2	5	16	"
1	35	6	13	15	All
2	36	5	12	16	"
3	35	6	5	23	"
4	37	4	8	20	"
5	38	3	5	23	"

Table 5-11

Pattern Classifier Results
Final Machine Configuration

Category One Data -- Measure Set C

Number of first stage prototype points:
Cough - 15 Artifact - 13

First stage - second layer decisions:

	C/C	A/C	C/A	A/A
Excluding Training sub-set	208	72	114	96
Training sub-set	134	30	43	89
Total	342	102	157	185

Second stage decisions:

Machine No.	C/C	A/C	C/A	A/A	Patterns Considered
1	8	2	5	2	Excl. training set
2	7	3	5	2	"
3	8	2	4	3	"
4	8	2	1	6	"
5	8	2	4	3	"
1	24	7	8	13	Training set
2	19	12	14	7	"
3	25	6	8	13	"
4	23	8	3	18	"
5	27	4	6	15	"
1	32	9	13	15	All
2	26	15	19	9	"
3	33	8	12	16	"
4	31	10	4	24	"
5	35	6	10	18	"

Table 5-12

Pattern Classifier Results
Final Machine Configuration

Category Two Data - Measure Set B

Number of first stage prototype points:
Cough - 15 Artifact - 15

First stage - second layer decisions:

	C/C	A/C	C/A	A/A
Excluding Training sub-set	247	115	205	439
Training sub-set	93	15	46	160
Total	340	130	251	599

Second stage decisions:

Machine No.	C/C	A/C	C/A	A/A	Patterns Considered
1	50	6	2	12	Excl. training set
2	44	12	3	11	"
3	31	25	2	12	"
4	46	10	1	13	"
5	37	19	5	9	"
1	50	6	2	12	Training set
2	36	20	5	9	"
3	37	19	2	12	"
4	53	3	2	12	"
5	48	8	2	12	"
1	100	12	4	24	All
2	80	32	8	20	"
3	68	44	4	24	"
4	99	13	3	25	"
5	85	27	7	21	"

Table 5-13

Pattern Classifier Results
Final Machine Configuration

Category Two Data - Measure Set C

Number of first stage prototype points:
Cough - 13 Artifact - 14

First stage - second layer decisions:

	C/C	A/C	C/A	A/A
Excluding Training sub-set	305	57	186	458
Training sub-set	96	12	20	186
Total	401	69	206	644

Second stage decisions:

Machine No.	C/C	A/C	C/A	A/A	Patterns Considered
1	51	5	1	13	Excl. training set
2	56	0	1	13	"
3	54	2	1	13	"
4	53	3	1	13	"
5	55	1	2	12	"
1	50	6	0	14	Training set
2	55	1	1	13	"
3	46	10	0	14	"
4	55	1	1	13	"
5	56	0	0	14	"
1	101	11	1	27	All
2	111	1	2	26	"
3	100	12	1	27	"
4	108	4	2	26	"
5	111	1	2	26	"

Per cent correct classifications (all patterns):

Machine No.	Cough	Artifact	Overall
1	90.1	96.4	91.4
2	99.1	92.8	97.9
3	89.3	96.4	90.7
4	96.4	92.8	95.7
5	99.1	92.8	97.9

Table 5-14

Pattern Classifier Results
Final Machine Configuration

Category Two Data - Measure Set D

Number of first stage prototype points:
Cough - 15 Artifact - 12

First stage - second layer decisions:

	C/C	A/C	C/A	A/A
Excluding Training sub-set	303	59	140	504
Training sub-set	98	10	24	182
Total	401	69	164	686

Second stage decisions:

Machine No.	C/C	A/C	C/A	A/A	Patterns Considered
1	49	7	1	13	Excl. training set
2	51	5	1	13	"
3	50	6	1	13	"
4	51	5	1	13	"
5	53	3	3	11	"
1	51	5	0	14	Training set
2	52	4	1	13	"
3	48	8	1	13	"
4	52	4	0	14	"
5	54	2	0	14	"
1	100	12	1	27	All
2	103	9	2	26	"
3	98	14	2	26	"
4	103	9	1	27	"
5	107	5	3	25	"

All of the preliminary machines investigated show a pronounced tendency to classify artifacts as coughs when applied to patterns not in the training set.

The second stage of the pattern classifier presumes statistical independence of the binary input. That this assumption is flagrantly violated with respect to machines 2 and 4 in the parallel hyperplane configuration is reflected in the classification results. It is noted that, although measure set B is of lower dimensionality than measure set A, there is no apparent degradation of classifier performance. This can probably be traced to approximately equal number of hyperplanes implemented by the first stage for both measure sets.

An overall evaluation of this configuration indicates that it is somewhat less than satisfactory for the data with which the research described by this dissertation is concerned.

Tables 5-5 and 5-6 are tabulations of the results obtained for a modified parallel hyperplane first stage configuration. The modification resulted in a cell defined by adjacent hyperplanes being deleted unless it was occupied by at least 1% of the training sub-segments.

It would be expected that classification results would be degraded because of the lower dimensionality of the binary vector presented to machines 2 and 4 of the second stage of the pattern classifier and because of the increased number of errors in the first stage-second layer decisions,

which would be reflected as an increased error rate for machines 1 and 3. That such was the case is apparent from a comparison of the classification results on the training set for machine 3.

Tables 5-7 and 5-8 are tabulations of the results achieved by the preliminary machine which uses as a first stage a minimum distance classifier with respect to point sets.

The prototype points are determined by assigning a training sub-segment to an existing cluster (for which the prototype points are centroids) if it is within a pre-specified distance from its centroid and then updating the mean value of the cluster. If no such cluster exists, the training sub-segment is used to define a new cluster centroid. This procedure was reiterated until all training sub-segments had been assigned to a cluster. Cough and artifact training sub-segments were considered separately. A maximum of twenty clusters was allowed for each of the classes. If training sub-segments existed for which clusters were not found after the maximum number of allowable clusters had been formed, the radius of the hypersphere which defined the cluster boundary was increased by a small fixed increment and the process repeated. This was continued until all training sub-segments had been assigned to a cluster.

Unless at least seven of the artifact training sub-segments or eight of the cough sub-segments fell within a cluster boundary, the centroid of the cluster was deleted from the set of prototype points. The results indicated that this requirement resulted in the deletion of too many

of the originally determined prototype points. Although improved results could have been expected had a modification to the above procedure been made, the configuration was so similar to that of the final configuration selected that only investigation of the latter was pursued.

The estimated a priori occurrence of a cough was 0.596. It was noted that in no case did this a priori probability change the decision which would have been made had a maximum likelihood criterion been used.

5.5.2 Final Machine Configuration Results

Tables 5-9 through 5-11 are tabulations of the results of machine training and recognition on category one data. Tables 5-12 through 5-14 display the results for category two data. As previously noted, for category two data, the machine decisions are maximum likelihood classifications.

The "minimum distance with regard to point sets" first stage configuration of this machine differed from that of the machine previously discussed only in the manner in which the prototype points were determined. As before, a maximum of twenty each artifact and cough clusters were allowed. Prior to training, these forty clusters were assigned arbitrarily valued centroids. During training, a sub-segment was assigned to the cluster within its class (cough or artifact) for which the distance between the sub-segment and the cluster centroid were closest (in the Euclidian sense) compared to the other clusters within the class. The centroid of the cluster so chosen was then updated to include the newly assigned sub-segment.

The distance of the cluster centroids from the origin were stored in central memory. When a centroid was updated, the change in its distance from the origin was computed. The training sub-set was presented repeatedly until the changes in distance from the centroids of all clusters to the origin were negligible, at which time first stage training was complete. Clusters which contained less than three of the training sub-segments were deleted.

A comparison of Tables 5-9 and 5-10 again indicates that the pattern set with the lowest dimensionality yields a higher percentage of correct decisions. This is probably due to the relative number of cough and artifact prototype points implemented by the first stage of the classifier. For the noisy category one signals, an increase in dimensionality of the pattern results in a greater percentage of pattern space being occupied by cough patterns with a resultant increase in the number of erroneous artifact classifications.

In Table 5-10 it will be noted that although the classification for patterns excluding the training set results are quite acceptable for Machines 3 and 5, the overall training results on the total data set indicate that if additional patterns were included in the data a higher error rate would result.

The results obtained by the use of the measures contained in measure set C (which was chosen on the basis of category two data)

for the classification of category one data are tabulated in Table 5-11. It was not expected that this set of measures would yield as high a percentage of correct classifications as measures chosen on the basis of category one data if the measure selection technique used was valid. It will be noted that such was the case.

The same argument applies to the case for which the results are tabulated in Table 5-12.

The training and classification results for category two data with respect to the measures chosen from a consideration of category two data are listed in Tables 5-13 and 5-14. As would normally be expected, measure set C (fifty dimensions) yielded fewer erroneous classifications than did measure set D (twenty-five dimensions).

The results indicate that second stage machines two, four and five appear to be more satisfactory than machines one and three. If the pattern classifier was to be realized by real time circuitry, machine 5 should be used since fewer calculations are required for this configuration. This would result in less complex circuitry.

CHAPTER VI

CONCLUSIONS

6.1 Summary and Recommendations

A model for a pattern classifier which yields usable results has been presented. The model circumvents the difficulties attached to estimating a conditional multi-dimensional probability function by effecting a preliminary piecewise linear partition of pattern space and assigning binary values to the resulting cells.

More satisfactory results were obtained for simulated data than for data recorded in the hospital environment. It was postulated that this was due to the presence of contaminating noise in the category one data. Use of the simulated data was justifiable since a real time machine would probably be placed in close proximity to the patient being monitored. This would result in higher quality signals being available at the input to the classifier.

A measure selection technique was examined. Experimental results indicate that the procedure is adequate for a preliminary selection of measures, but that a final selection of a set of measures should be made on an empirical basis.

The measures selected for use by the various classifiers are satisfactory for either real time or general purpose computer recognition. In the latter case, however, analog pre-processing to obtain a representation

of the spectral measures is appropriate due to the disproportionate calculation time required to compute these measures. Segmentation and thresholding of the signals should be implemented prior to digitization of the data.

If the present recording procedure is continued, an effort should be made to improve the quality of the resulting signal.

Experimental results indicate that the assumption of Markoff-1 dependency between the data sub-segments is a reasonable hypothesis. The configuration based on this assumption (machine number 5) may be implemented more simply by real time circuitry and requires fewer recognition computations by a general purpose machine than other machines with comparable results. It should therefore be chosen for the realization of the recognition process. This conclusion would not necessarily be valid if the pattern classifier was applied to data originating from an experiment other than that considered in this research.

6.2 Application Extensions

The model presented in this paper is suitable for use (with slight modification) with numerous types of data, the primary requirement being that sequential segments of data from the same class be available.

An application directly related to that described by this paper would be the classification of coughs as having originated from a patient suffering from irreversible lung damage as contrasted with coughs emanating from a respiratory system in which the cough causing factors are of a temporary or reversible nature. The physiological model indicates that a

difference in the audible characteristics of the coughs (and possibly forced expirations) probably exists.

A study of the statistical characteristics of the measures calculated and their relationship to the physiological model would be rewarding. Additionally, such a study would indicate the degree of correlation between the various measures. If two or more measures are highly correlated statistically, only one should be used in the recognition process.

REFERENCES

1. Bickerman, H., and S. Itkin, "The Effect of a New Bronchodilator Aerosol on the Air Flow Dynamics of the Maximal Voluntary Cough of Patients with Bronchial Asthma and Pulmonary Emphysema," J. Chron. Dis., Vol. 8, No. 5, pp. 629-636, November, 1958.

2. Blackman, R., and J. Tukey, The Measurement of Power Spectra, Dover Publications, Inc., New York, 1959.

3. Chu, J. T., and J. C. Chueh, "Error Probability in Decision Functions for Character Recognition," Journal of the Association for Computing Machinery, Vol. 14, No. 2, pp. 273-280, April, 1967.

4. Cooley, James W., and J. W. Tukey, "An Algorithm for the Machine Calculation of Fourier Series," Math. Computation, Vol. 19, pp. 297-301, April, 1965.

5. Cooley, W. W., and P. R. Lohnes, Multivariate Procedures for the Behavioral Sciences, John Wiley and Sons, Inc., New York, 1962.

6. Deutch, Ralph, Estimation Theory, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1965.

7. DiRenzo, S., "Bronchial Dynamism," Radiology, Vol. 53, pp. 168-183, August 1949.

8. Fatechand, R., "Machine Recognition of Spoken Words," Advances in Computers, F. L. Alt, Ed., Academic Press, New York, 1960, pp. 193-227.

9. Hadley, G., Linear Algebra, Addison-Wesley, Reading, Mass., 1961.

10. Nilsson, Nils J., Learning Machines, McGraw-Hill, New York, 1965.

11. Papoulis, Athanasios, Probability, Random Variables, and Stochastic Processes, McGraw-Hill, New York, 1965.

12. Sebestyen, George S., and Edie, "An Algorithm for Non-Parametric Pattern Recognition," IEEE Trans. Elect. Comp., Vol. EC-15, No. 6, 1967.

13. Sebestyen, George S., Decision-Making Processes in Pattern Recognition, The Macmillan Co., New York, 1962.

14. Whittenberger and Mead, "Respiratory Dynamics During Cough," Tr. Nat. Tuberc. A., Vol. 48, pp. 414-418, 1952.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) The University of Texas Laboratories for Electronics & Related Science Research Austin, Texas 78712	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP
3. REPORT TITLE PATTERN RECOGNITION APPLIED TO COUGH CATEGORIZATION	
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Scientific Interim	
5. AUTHOR(S) (First name, middle initial, last name) Joseph L. Devine, Jr. and Ashley J. Welch	
6. REPORT DATE September 15, 1967	7a. TOTAL NO. OF PAGES 118 7b. NO. OF REFS 14
8a. CONTRACT OR GRANT NO. AF-AFOSR-766-67 b. PROJECT NO. 4751 c. 61445014 d. 681305	9a. ORIGINATOR'S REPORT NUMBER(S) JSEP, Technical Report No. 40 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFOSR-67-2137
10. DISTRIBUTION STATEMENT 1. Distribution of this document is unlimited.	
11. SUPPLEMENTARY NOTES TECH, OTHER	12. SPONSORING MILITARY ACTIVITY JSEP through AF Office of Scientific Research (SREE) 1400 Wilson Boulevard Arlington, Virginia 22209

13. ABSTRACT

The particular problem with which the research was concerned was the development of a technique to discriminate between coughs and other audible phenomena which originate in a hospital environment. Pattern recognition provided such a technique. Experimental data was available in the form of audio tape recordings.

Implementation of a Bayes categorization decision requires knowledge of the underlying conditional joint probability density functions of the measures which typify the patterns to be recognized. An adaptive pattern classifier model was presented which circumvented the difficulty of estimating these functions. The model is generally applicable to the two-class case in which the patterns to be classified consist of sequential segments of data known to have originated from the same class. The model took the form of a layered machine. The first stage was a minimum distance classifier with respect to point sets while the second stage utilized the first stage binary valued outputs to implement a Bayes decision.

The feature extraction and measure selection problems were examined experimentally. Feature calculation algorithms were developed which are generally applicable to time varying signals. The effectiveness of an algorithm for selection of a set of candidate measures was verified.

Experimental results indicated that, for the particular data with which the research was concerned, an assumption of Markoff-1 dependency between sequential first stage decisions of the pattern classifier was warranted. A pattern classifier which was based on this assumption classified 97.9% of the patterns presented to its input correctly.

DD FORM 1473

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
PATTERN RECOGNITION ADAPTIVE CLASSIFICATION BIOMEDICAL SIGNAL PROCESSING						